CORE
Metadata, citation and similar papers at core.ac.uk
Provided by ASU Digital Repository

A Power Study of Gffit Statistics as Components of Pearson Chi-Square

by

Junfei Zhu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2017 by the
Graduate Supervisory Committee:

Mark Reiser, Chair
John Stufken
Yi Zheng
Robert St Louis
Ming-Hung Kao

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

The Pearson and likelihood ratio statistics are commonly used to test goodness-of-fit for models applied to data from a multinomial distribution. When data are from a table formed by cross-classification of a large number of variables, the common statistics may have low power and inaccurate Type I error level due to sparseness in the cells of the table. The GFfit statistic can be used to examine model fit in subtables. It is proposed to assess model fit by using a new version of GFfit statistic based on orthogonal components of Pearson chi-square as a diagnostic to examine the fit on two-way subtables. However, due to variables with a large number of categories and small sample size, even the GFfit statistic may have low power and inaccurate Type I error level due to sparseness in the two-way subtable. In this dissertation, the theoretical power and empirical power of the GFfit statistic are studied. A method based on subsets of orthogonal components for the GFfit statistic on the subtables is developed to improve the performance of the GFfit statistic. Simulation results for power and type I error rate for several different cases along with comparisons to other diagnostics are presented.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

Table Page

Table Page

Table Page

Table Page

,

CHAPTER 1

INTRODUCTION

The cross classification of several categorical variables produces a large contingency table. If a model is fit to the table, usually the Pearson chi-square and the likelihood ratio statistics are used to evaluate the goodness of fit. Suppose we have q categorical variables and the i-th variable has c_i categories. Thus there are $k = \prod_{i=1}^{q} c_i$ cells, also called response patterns in the cross-classified table. Then f_r is the sample proportion of the r-th response pattern and $\hat{\pi}_r$ is the estimated probability of the r-th response pattern. The Pearson chi-square(χ^2) and the likelihood ratio(LR) statistics are defined as follows:

$$LR = 2n \sum_{r=1}^{k} f_r \ln\left(\frac{f_r}{\hat{\pi}_r}\right)$$

$$\chi^2 = n \sum_{r=1}^{k} \frac{(f_r - \hat{\pi}_r)^2}{\hat{\pi}_r}$$

If the number of observations in each response pattern is large enough and under the conditions (Koehler and Larntz, 1980) that i) $H_0: \pi = \pi(\theta)$, ii) k is fixed and iii) $\min_{1 \le r \le k} n\pi_r \to \infty$ for $n \to \infty$, both Pearson chi-square and likelihood ratio statistics are approximately distributed chi-square with degree of freedom equal to $k - 1 -$ number of estimated parameters.

However, in presence of sparse data, these two statistics may not follow the chi-square distribution even if the sample size is large. When the ratio of the sample size to the number of cells is relatively small, contingency tables are said to be sparse (Agresti & Yang, 1987), but sparseness can also be produced by very skewed cell frequencies in some cases. There is no universal agreement on what constitutes a small expected

frequency. Cochran (1954) suggested that most expected frequencies should be at least five. Cramer (1946) has suggested 10 and Kendall (1952) has suggested 20.

One way to solve this problem is to use asymptotic normality of the Pearson and likelihood ratio statistics when both the sample size and number of cells become large. Morris (1975) showed that both the Pearson's chi-square statistic and likelihood ratio statistic have asymptotic normal distributions under certain conditions. Koehler and Larntz (1980) suggest that because of the different influence of very small observed counts on Pearson's chi-square statistic and likelihood ratio statistic, the asymptotic means and variances of these two statistics are different. Koehler and Larntz (1986) also provides a Monte Carlo study of these two statistics for loglinear models. The results show that generally the normal approximation is more accurate for likelihood ratio statistic than for Pearson's chi-square statistic.

Another way to solve the sparseness problem is to use statistics based on the marginal frequencies. There are several statistics of this kind, such as Maydeu-Olivares' statistic(2005), Bartholomew's Y statistic(2002) and the orthogonal components of Pearson chi-square. Reiser (2008) introduced a score statistic based on the overlapping cells that correspond to the second-order marginal frequencies. Then orthogonal components of the Pearson-Fisher statistic are defined on marginal frequencies. The score statistic is shown to be a sum of these orthogonal components and is denoted $X^2_{[2]}$. $X^2_{[2]}$, the Y statistic and Maydeu-Olivares' statistic are all score statistics and distribute asymptotically chi-square. The marginal frequencies are just linear combinations of the joint frequencies. Thus we can introduce a matrix **H**, which I will define later, to compute

the marginal frequencies. Since all the statistics mentioned here are based on the marginal frequencies, they can be computed easily using the **H** matrix.

When using $X^2_{[2]}$ to test the goodness-of-fit of a model, it may have higher power for certain alternative hypotheses because it represents a test that is "focused" on the second-order marginal. If lack of fit is present in second-order marginal, then $X^2_{[2]}$ would have higher power than an omnibus statistic such as the Pearson chi-square. But if lack of fit is present in higher-order marginal, then $X^2_{[2]}$ may have lower power. Similarly, since $X^2_{[2]}$ is the sum of all the orthogonal components, it can be considered as an omnibus statistic on the second-order marginals. If we just sum up a subset of these orthogonal components corresponding to variable i and variable j, then we get a statistic only focused on variable i and variable j on the second-order marginal. This statistic is denoted $GFfit^{(ij)}_{\perp}$ and it is distributed asymptotically chi-square. $GFfit^{(ij)}_{\perp}$ can be used as a diagnostic to detect the source of lack of fit when the model does not fit the observed data. If lack of fit is present in the association between variable i and variable j, then $GFfit^{(ij)}_{\perp}$ would have higher power than an omnibus statistic on the second-order marginals such as $X^2_{[2]}$.

My research is focused on $X^2_{[2]}$ and $GFfit^{(ij)}_{\perp}$. I studied their Type I error rate and power and compared their performance with several other statistics. Monte Carlo simulations were conducted to study the empirical type I error rate and power of these statistics. Theoretical power calculation were also conducted for $X^2_{[2]}$ and $GFfit^{(ij)}_{\perp}$. Besides these type I error and power studies, I also improved the $GFfit^{(ij)}_{\perp}$. Although $GFfit^{(ij)}_{\perp}$ is a good remedy to the problem of sparseness because it is calculated from marginal two-

way tables, sometimes even $GFfit_{\perp}^{(ij)}$ may have low power and inaccurate Type I error level due to severe sparseness in a two-way subtable when the number of categories is large and response variables have a skewed distribution. In that case, the distribution of $GFfit_{\perp}^{(ij)}$ may not be well approximated by the chi-square distribution even if the total sample size is large. I improved $GFfit_{\perp}^{(ij)}$ in the sparse case by selecting a subset of orthogonal components chosen systematically to reduce the impact of sparseness. I denote this improved statistic $GFfit_{\perp(\mathrm{t})}^{(ij)}$ where t represents the t orthogonal components chosen to compute this statistic. $GFfit_{\perp(\mathrm{t})}^{(ij)}$ is distributed asymptotically chi-square with t degrees of freedom. The same idea was applied to $X_{[2]}^2$ and I denote the new statistic $X_{[2][t]}^2$. Type I error rate and power were studied for $GFfit_{\perp(\mathrm{t})}^{(ij)}$ and $X_{[2][t]}^2$.

In this dissertation, I will present a literature review first. In the literature review, I will first introduce the traditional Pearson chi-square statistic and the problem of sparseness. Then I will introduce the orthogonal components along with several other statistics based on marginal proportions. After the literature review, I will show theoretical and empirical studies of the GFfit statistic. Simulation results and an application will be presented. Finally, summary and discussion will be presented.

CHAPTER 2

LITERATURE REVIEW

II.1.1 Pearson Chi-square Statistic

Pearson's chi-square test was the first goodness of fit test and perhaps one of the most frequently used statistical tests. So in this section I will first review Pearson's chi-square test. Suppose there is a multinomial distribution involving k cells with known cell probabilities$[\pi_1, \pi_2, \ldots, \pi_k], 0 < \pi_i < 1; \quad \sum \pi_i = 1$. For a random sample of fixed size n, let $[f_1, f_2, \ldots, f_k]$, $0 \le f_i \le n$; $\sum f_i = n$ be the random frequency counts in the respective cells $\{1, 2, \ldots, k\}$. It follows that

$$E[f_i] = n\pi_i = E_i;$$

$$Var(f_i) = n\pi_i (1 - \pi_i);$$

$$Covar (f_i, f_j) = -n\pi_i\pi_j; \; i \neq j$$

Since $\sum f_i = n, Var(\sum f_i) = 0$ and hence the joint distribution of $[f_1, f_2, \ldots, f_k]$is singular. Denote the $k \times k$ matrix of the variances and covariances of the f's by W. Note that

$$W = n[\pi^{\delta} - \pi\pi']$$

where

$$\pi = (\pi_1, \pi_2, \ldots, \pi_k)'$$

$$\pi^{\delta} = \text{diag}\{\pi\text{'s}\} = \begin{bmatrix} \pi_1 & 0 & \ldots & 0 \\ 0 & \pi_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \pi_k \end{bmatrix}$$

It can be proven that

$$\text{Rank } (W) = k - 1$$

Then the Pearson chi-square statistic is defined as

$$\chi^2 = \sum_{i=0}^{k} \frac{(f_i - E_i)^2}{E_i}$$

It follows that when n tends to infinity, the limiting distribution of χ^2 is that of a Chi-square distribution with k-1 df. To prove this, for every single selected random sampling unit, we can introduce the 'indicator functions' of the k cells. That means, for the ith sampled unit,

$$\{I_i(1), \dots, I_i(\mathrm{t}), \dots, I_i(\mathrm{k})\}$$

denote the underlying random indicator functions

$$I_i(\mathrm{t}) = \begin{cases} 1 & if\ ith\ sampled\ unit\ belongs\ to\ cell\ t \\ 0 & otherwise \end{cases}$$

This is true for t = 1, …, k; i =1, 2, …, n.

Clearly,

$$\sum_{t=1}^{k} I_i(\mathrm{t}) = 1\ for\ each\ i = 1, 2, \dots, n$$

$$\sum_{i=1}^{n} I_i(\mathrm{t}) = f_t\ for\ each\ t = 1, \dots, k$$

Therefore for every i = 1, 2, …, n, $\{I_i(1), \dots, I_i(\mathrm{t}), \dots, I_i(\mathrm{k})\}$ follows a singular multinomial distribution with parameters $[\pi_1, \pi_2, \dots, \pi_k]$ and are i.i.d. By the central limit theorem, the sample average counts

$$[\frac{f_1}{n}, \dots, \frac{f_k}{n}]$$

follow a singular multivariate normal distribution with rank k-1 since the variance-covariance matrix W of the f_i's has rank k-1. Therefore

$$\mathrm{Q} = (f - n\pi)' W^{+} (f - n\pi)$$

has chi-square distribution with k-1 df, where W^+ is the Moore-Penrose g-inverse of the W and

$$W^+ = \frac{1}{n}\left[\pi^{-\delta} - 11'\right]$$

$\pi^{-\delta}$ is the inverse of π^{δ} and 1 is a column vector of 1's.

Thus

$$\mathrm{Q} = (\mathrm{f}-\mathrm{E})'W^+(f-E) = (\mathrm{f}-\mathrm{E})'\frac{1}{n}\left[\pi^{-\delta} - 11'\right](f-E)$$

$$= (\mathrm{f}-\mathrm{E})'\left[\frac{1}{n}\pi^{-\delta} - \frac{1}{n}11'\right](f-E) = (\mathrm{f}-\mathrm{E})'\left[E^{-\delta} - \frac{1}{n}11'\right](f-E)$$

$$= (\mathrm{f}-\mathrm{E})'E^{-\delta}(f-E) - \frac{1}{n}(\mathrm{f}-\mathrm{E})'11'(f-E)$$

$$= \sum_{i=1}^{k}\frac{(f_i-E_i)^2}{E_i} - \frac{1}{n}\sum_{i=1}^{k}(f_i-E_i)\,1'(f-E)$$

$$= \sum_{i=1}^{k}\frac{(f_i-E_i)^2}{E_i} - \frac{1}{n}\times 0\times 1'(f-E) = \sum_{i=1}^{k}\frac{(f_i-E_i)^2}{E_i}$$

Therefore we have shown that the limiting distribution of χ^2 is that of a Chi-square with k-1 degrees of freedom.

If the cell probabilities $[\pi_1, \pi_2, \dots, \pi_k]$ are unknown, we will use a model with g unknown parameters to estimate the cell probabilities. To estimate the probabilities, we need to estimate the unknown parameters from the sample first. Then we replace the expected frequencies E's by estimated frequencies $\hat{E}'s$, where

$$\hat{E}_i = n\hat{\pi}_i\,; i = 1,2,\dots,k$$

Then we compute the Pearson chi-square statistic as

$$\chi^2_{PF} = \sum_{i=0}^{k} \frac{(f_i - \hat{E}_i)^2}{\hat{E}_i}$$

Since the g unknown parameters have to be estimated from the data, we lose some degrees of freedom. Fisher (1924) gives the first derivation of the correct degrees of freedom, namely $k - g - 1$ when g parameters are estimated from the data. Thus this statistic is also called the Pearson-Fisher statistic.

II.1.2 The Partition of χ^2

Lancaster (1969) introduced the partition of χ^2. In the earlier section, we suppose that n observations are given on a set of the k indicator variables of the multinomial distribution. If a subset of $(k-1)$ indicator variables is chosen, the remaining variable is determined since only $(k-1)$ of these indicator variables are linearly independent. Equivalently, any set of $(k-1)$ orthonormal functions, $\{U^{*(i)}\}$, may be considered and their standardized sums,

$$U^{(i)} = n^{-\frac{1}{2}} \sum_{j=1}^{n} U_j^{*(i)}$$

Then

$$\chi^2 = \sum_{i=1}^{k-1} U^{(i)}$$

And χ^2 is invariant for any choice of the set $\{U^{*(i)}\}$.

II.1.3 Score Statistics

Suppose we have a random sample $X_1, X_2, \dots, X_n$ from a continuous distribution with pdf $f(x;\theta)$ where

$$\theta = (\theta_1, \dots, \theta_k)' \in \Theta, \text{the parameter space}$$

We want to test the null hypothesis $H_0: \theta = \theta_0$ against $H_a: \theta \neq \theta_0$. Note that if the distribution of X is discrete, the following results still hold. Let L be the likelihood function, then

$$L = \prod_{i=1}^{n} f(x_i; \theta)$$

$$\text{the score } U(\theta) = \left(U_i(\theta)\right), \text{where } U_i(\theta) = \frac{\partial log L}{\partial \theta_i}$$

$$\text{the information matrix } I(\theta) = \left(I_{ij}(\theta)\right), \text{where } I_{ij}(\theta) = E_\theta\left[U_i(\theta)U_j(\theta)\right]$$

$$= -E_\theta\left[\frac{\partial^2 log L}{\partial \theta_i \partial \theta_j}\right]$$

Then the score test statistic is defined as

$$S = \{U(\theta_0)\}'\{I_{ij}(\theta_0)\}^{-1}\{U(\theta_0)\}$$

Under the null hypothesis, S is asymptotically distributed chi-square with k degrees of freedom, where k is the number of elements in θ, or equivalently, the dimension of the parameter space.

However, sometimes we are only interested in several particular parameters. For example, when testing for a normal mean, usually the variance is also unknown but we are not interested in it. In this case the unknown variance will enter the problem as a 'nuisance' parameter. To deal with this problem, we let $f(x; \gamma)$ be the probability density function and γ is the parameter vector. γ can be partitioned into

$$\gamma = (\theta', \beta')'$$

where θ is a k×1 vector of real parameters and β is a q×1 vector of nuisance parameters. Then we can partition the score and information matrix into

$$\mathrm{U} = \mathrm{U}(\gamma) = \begin{pmatrix} U_\theta(\gamma) \\ U_\beta(\gamma) \end{pmatrix}$$

$$\mathrm{I} = \begin{pmatrix} I_{\theta\theta} & I_{\theta\beta} \\ I_{\beta\theta} & I_{\beta\beta} \end{pmatrix}$$

Then $\Sigma(\gamma)$ is defined by

$$\Sigma(\gamma) = I_{\theta\theta}(\gamma) - I_{\theta\beta}(\gamma)\{I_{\beta\beta}(\gamma)\}^{-1} I_{\beta\theta}(\gamma)$$

Cox and Hinkley (1974, Section 9.3) showed that $\{\Sigma(\gamma)\}^{-1}$ is the asymptotic covariance matrix of $\hat{\theta}$ and $\Sigma(\gamma)$ is the asymptotic covariance matrix of $U_\theta(\gamma)$.

Then the score test statistic is defined as

$$\hat{\mathrm{S}} = \{U_\theta(\hat{\gamma}_0)\}'\{\Sigma(\hat{\gamma}_0)\}^{-1}\{U_\theta(\hat{\gamma}_0)\}$$

where $\hat{\gamma}_0$ is the ML estimator of γ under the null hypothesis H_0: $\theta = \theta_0$, in which θ is restricted to taking the value θ_0. Under the null hypothesis, $\hat{\mathrm{S}}$ is asymptotically distributed chi-square with k degrees of freedom.

II.1.4 Components

Suppose we want to test the null hypothesis that m cell probabilities are $p_j = p_j(\beta), j = 1, \ldots, m$. An alternative is to take the order k 'smooth' probability function

$$\pi_j(\theta, \beta) = C(\theta, \beta) \exp\left\{\sum_{i=1}^{k} \theta_i h_{ij}(\beta)\right\} p_j(\beta), \quad j = 1, \ldots., m$$

where θ is a k×1 vector of real parameters, β is a q×1 vector of nuisance parameters and $C(\theta, \beta)$ is a normalizing constant that ensures $\sum_{j=1}^{m} \pi_j(\theta, \beta) = 1$. For each i, $i = 1, \ldots, k$, the $h_{ij}(\beta)$ are values taken by a random variable H_i with $P\left(H_i = h_{ij}(\beta)\right) = \pi_j(\theta, \beta), j = 1, \ldots, m$. Here $k \leq m - 1$ since the parameter space has dimension m-1. Testing H_0: $\pi_j(\theta, \beta) = p_j(\beta)$ vs H_a: $\pi_j(\theta, \beta) \neq p_j(\beta)$ is equivalent to test H_0: $\theta =$

$0\ vs\ H_a: \theta \neq 0$. For convenience, we drop the argument θ and β from $\pi_j(\theta, \beta), p_j(\beta)$ and $h_{ij}(\beta)$.

Suppose a random sample of n observations is taken and the number of observations in the jth class is $N_j, j = 1, \ldots, m$. Write H $= (h_{ij})$, $N = (N_j)$ and $p = (p_j)$

Rayner and Best(1989) have shown that the Pearson-Fisher statistics ${\chi^2}_{PF}$ can be partitioned into components via ${\chi^2}_{PF} = \hat{V}_1^2 + \cdots + \hat{V}_{m-q-1}^2$ in which the $\hat{V}_r$ are asymptotically standard normal and asymptotically independent, being defined by $\hat{V}_r = \sum_{j=1}^{m} \hat{h}_{rj} N_j / \sqrt{n}$.

II.2 Problem of Sparseness

Suppose we have q categorical variables and the i-th variable has c_i categories. Thus there are $k = \prod_{i=1}^{q} c_i$ cells, also called response patterns in the cross-classified table. When the sample size to the number of cells is relatively small, contingency tables are said to be sparse (Agresti & Yang, 1987). When there is a problem of sparseness, a test statistic based on an asymptotic chi-square distribution may no longer follow a chi-square distribution. There is no universal agreement on what constitutes a small expected frequency. Cochran (1954) suggested that most expected frequencies should be at least five. Cramer (1946) has suggested 10 and Kendall (1952) has suggested 20. When sparseness is present in a set of frequencies, combining cells or adding a small constant such as 0.5 to each cell are sometimes attempted (Goodman, 1964).

One way to solve the problem of sparseness is to consider other distribution for the goodness-of-fit statistics. If the number of observations in each response pattern is large enough and under the conditions that $i)\ H_0: \pi = \pi(\theta),\ ii)\ k$ is fixed and

$iii)$ $min_{1\leq r\leq k} n\pi_r \to \infty$ for $n \to \infty$, the Pearson's chi-square statistic is distributed asymptotically chi-square. Morris (1975) showed that both the Pearson's chi-square statistic and likelihood ratio statistic have asymptotic normal distributions under conditions that allow both n and k to become large without necessarily requiring that $min_{1\leq r\leq k} n\pi_r \to \infty$, which means the number of cells is increased when the sample size is increased. Consider the sequence of multinomial random vectors

$$\{(N_{1,k(i)}, N_{2,k(i)}, \dots, N_{k(i),k(i)})\}_{i=1}^{\infty}$$

where the i-th vector in the sequence has $k(i)$ cells. The sample size is $n_k = \sum_{j=1}^{k} N_{j,k}$ and the probability vector is $(p_{1k}, p_{2k}, \dots, p_{kk})$ with $\sum_{j=1}^{k} p_{jk} = 1$. Morris (1975) derived a central limit theorem for the Pearson statistic. Under the null hypothesis, the sufficient conditions for asymptotic normality as k $\to \infty$ are (a) $max_{1<i<k} p_{ik} = o(1)$ as k $\to \infty$ and (b) $n_k p_{ik}$ is uniformly bounded below by some constant. When the null hypothesis is true, the asymptotic mean and variance for the Pearson statistic are

$$\mu_{P,k} = k$$

$$\sigma_{P,k}^2 = 2\mathrm{k} + \sum_{j=1}^{k} \frac{(1 - k^{-1} p_{jk})}{n_k p_{jk}}$$

Note that when the expected frequencies are not all equal, $\sigma_{P,k}^2$ may be much larger than the chi-square variance on k $-$ 1 degrees of freedom.

Koehler and Larntz (1980) suggest that because of the different influence of very small observed counts on Pearson's chi-square statistic and likelihood ratio statistic, the asymptotic means and variances of these two statistics are different. Koehler and Larntz (1986) also provides a Monte Carlo study of these two statistics for loglinear models. The

results show that generally the normal approximation is more accurate for likelihood ratio statistic than for Pearson's chi-square statistic.

Another way to solve the problem of sparseness is to use statistics based on the marginal frequencies. There are several statistics of this kind, which will be introduced later.

II.3 Orthogonal Components Based on Marginal Proportions

Reiser (2008) introduced a score statistic based on the overlapping cells that correspond to the first and second-order marginal frequencies. Then orthogonal components of the Pearson-Fisher statistic are defined on marginal frequencies. The score statistics is shown to be a sum of these orthogonal components

II.3.1 First- and Second-order Marginals.

The relationship between joint proportion and first- and second-order marginal cane be shown by using zeros and ones to code the levels of categorical response variables, $Y_i, i = 1, 2, \ldots q$. Each Y_i has $c \geq 2$ categories. A specific cell from the contingency table, sometimes called a response pattern, can be indicated by a $(c-1)q$-dimensional vector of zeros and ones. Then a $T = c^q$-dimensional set of response patterns can be generated by varying the levels of the q^{th} variable most rapidly, the $q^{th} - 1$ variable next, etc. Define $\boldsymbol{V}$ as the T by $(c-1)q$ matrix with response patterns as rows.

For $q = 3$ and $c = 2$,

$$\boldsymbol{V} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

For $q = 3$ and $c = 3$, $\boldsymbol{V}$ is a 27 by 6 matrix:

$$\boldsymbol{V} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

The matrix $\boldsymbol{V}$ can be generated by kernel patterns. $\boldsymbol{V}$ has $(c-1)$ kernel patterns, each of dimension c. In general, the i^{th} kernel pattern, $f_i = (0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0)'$ with the 1 on the (i+1)-th position and $i = 1, 2, \ldots, c-1$. For $c = 2$, the kernel pattern is $f_1 = (0 \quad 1)'$, and for $c = 3$, the kernel patterns are $f_1 = (0 \quad 1 \quad 0)'$ and $f_2 = (0 \quad 0 \quad 1)'$. The matrix $\boldsymbol{V}$ can be generated by Kronecker products of the kernel patterns with the vector $\mathbf{1}_c$, which is a vector of length c where each element is 1. The patterns of columns are $(c-1)$ columns $f_i \otimes (\mathbf{1}_c \otimes \mathbf{1}_c \otimes \cdots \otimes \mathbf{1}_c), i = 1, \ldots, c-1$, followed by $(c-1)$ columns $\mathbf{1}_c \otimes f_i \otimes (\mathbf{1}_c \otimes \cdots \otimes \mathbf{1}_c), i = 1, \ldots, c-1$, continuing until $(c-1)$ columns $(\mathbf{1}_c \otimes \mathbf{1}_c \otimes \cdots \otimes \mathbf{1}_c) \otimes f_i, i = 1, \ldots, c-1$.

With $q = 3$ and $c = 2$,

$$\boldsymbol{V} = (f_1 \otimes (\mathbf{1}_2 \otimes \mathbf{1}_2), \mathbf{1}_2 \otimes (f_1 \otimes \mathbf{1}_2), (\mathbf{1}_2 \otimes \mathbf{1}_2) \otimes f_1)$$

With $q = 3$ and $c = 3$,

$$\begin{matrix} \boldsymbol{V} = (f_1 \otimes (\mathbf{1}_3 \otimes \mathbf{1}_3), & f_2 \otimes (\mathbf{1}_3 \otimes \mathbf{1}_3), & \mathbf{1}_3 \otimes (f_1 \otimes \mathbf{1}_3), \\ \mathbf{1}_3 \otimes (f_2 \otimes \mathbf{1}_3), & (\mathbf{1}_3 \otimes \mathbf{1}_3) \otimes f_1, & (\mathbf{1}_3 \otimes \mathbf{1}_3) \otimes f_1) \end{matrix}$$

Define $\mathbf{H}_{[1]} = \boldsymbol{V}'$, where h_{ls} is an element of $\mathbf{H}_{[1]}, l = 1, 2, \dots, q(c-1), s = 1, 2, \dots, T$. Then, under some specific model $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$, which we will introduce later, the first-order marginal proportion for variable Y_i can be defined as

$$\pi^{(i)}(a;\boldsymbol{\theta}) = \text{Prob}(Y_i = a|\boldsymbol{\theta}) = \sum_s h_{ls}\pi_s(\boldsymbol{\theta}) = \boldsymbol{h}_{\boldsymbol{l}}'\boldsymbol{\pi}(\boldsymbol{\theta}), a = 2, \dots, c, l = (c-1)(i-1) + a - 1,$$

Where $\boldsymbol{h}_{\boldsymbol{l}}'$ is row l of matrix $\mathbf{H}_{[1]}$. Then the true first-order marginal proportion is given by

$$\pi^{(i)}(a) = \text{Prob}(Y_i = a) = \sum_s h_{ls}\pi_s = \boldsymbol{h}_{\boldsymbol{l}}'\boldsymbol{\pi}.$$

Under the model, the second-order marginal proportion for variable Y_i and Y_j can be defined as

$$\pi^{(ij)}(a,b;\boldsymbol{\theta}) = \text{Prob}\left(Y_i = a, Y_j = b\middle|\boldsymbol{\theta}\right) = \sum_s h_{ks}h_{ls}\pi_s(\boldsymbol{\theta}) = (\boldsymbol{h}_{\boldsymbol{k}}' \circ \boldsymbol{h}_{\boldsymbol{k}}')\boldsymbol{\pi}(\boldsymbol{\theta}),$$

Where $i = 1, \dots, q-1; j = i, \dots, q;\ k = (c-1)(i-1) + a - 1;\ l = (c-1)(j-1) + b - 1; a = 2, \dots, c; a = 2, \dots, c$; and $\boldsymbol{h}_{\boldsymbol{k}}' \circ \boldsymbol{h}_{\boldsymbol{k}}'$ represents the Hadamard product of rows k and l. Then the true second-order marginal proportion is given by

$$\pi^{(ij)}(a,b) = \text{Prob}(Y_i = a, Y_j = b) = \sum_s h_{ks}h_{ls}\pi_s = (\boldsymbol{h}_{\boldsymbol{k}}' \circ \boldsymbol{h}_{\boldsymbol{k}}')\boldsymbol{\pi},$$

The summation across the frequencies associated with the response patterns to obtain the marginal proportions represents a linear transformation of the frequencies in the multinomial vector $\boldsymbol{\pi}$ which can be implemented via multiplication by a certain matrix, denoted generally by $\mathbf{H}$. The symbol $\mathbf{H}_{[t]}$ denotes the transformation matrix that would produce marginal of order t. The symbol $\mathbf{H}_{[t:u]}, t \le u \le q$, denotes the transformation matrix that would produce marginal from order t up to and including order u.

For second-order marginal proportions, the rows of $\mathbf{H}_{[2]}$ are Hadamard products among the columns of V. For $q = 3$ and $c = 2$,

$$\mathbf{H}_{[2]} = \begin{bmatrix} (v_1 \circ v_2)' \\ (v_1 \circ v_3)' \\ (v_2 \circ v_3)' \end{bmatrix}$$

where v_i is the column i of matrix $\boldsymbol{V}$, and $v_i \circ v_j$ is the Hadamard product of columns i and j.

For $q = 3$ and $c = 3$, $\mathbf{H}_{[2]}$ is an 12 by 27 matrix:

$$\mathbf{H}_{[2]} = \begin{bmatrix} (v_1 \circ v_3)' \\ (v_1 \circ v_4)' \\ \vdots \\ (v_1 \circ v_5)' \\ (v_1 \circ v_6)' \\ \vdots \\ (v_3 \circ v_5)' \\ \vdots \\ (v_{i(c-1)} \circ v_{j(c-1)})' \end{bmatrix}$$

II.3.2 Higher-order Marginals

Matrix **H** for higher-order marginal can be defined in a similar way using Hadamard products among columns of $\boldsymbol{V}$. The third-order marginal proportions for variables $Y_i, Y_j \; and \; Y_k$ can be obtained by employing the matrix $\mathbf{H}_{[3]}$. Then we define

$$\mathbf{H}_{[t:u]} = \begin{bmatrix} \mathbf{H}_{[t]} \\ \mathbf{H}_{[t+1]} \\ \vdots \\ \mathbf{H}_{[u]} \end{bmatrix}$$

II.3.3 $\boldsymbol{X^2_{[t:u]}}$ Statistic

Now our null hypothesis is H_0: $\mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ and the test statistic is

$$X^2_{[t:u]} = \boldsymbol{e}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1}\boldsymbol{e}$$

$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = n^{-1}\boldsymbol{\Omega}_{\boldsymbol{e}}$ with $\boldsymbol{\Omega}_{\boldsymbol{e}}$ evaluated at the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$, and where

$$\boldsymbol{\Omega_e} = \mathbf{H}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')\mathbf{H}'$$

$D(\boldsymbol{\pi})$ =diagnal matrix with (s, s) element equal to $\pi_s(\boldsymbol{\theta})$

$$\mathbf{A} = D(\pi)^{-1/2}\frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

$$\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

$\boldsymbol{e} = \mathbf{H}(\boldsymbol{p} - \boldsymbol{\pi})$ is the matrix form of the marginal residuals

and $\boldsymbol{p}$ is the observed proportion

$\mathbf{H} = \mathbf{H}_{[1:2]}$ produces $X^2_{[1:2]}$ and $\mathbf{H} = \mathbf{H}_{[2]}$ produces $X^2_{[2]}$. It has been proved that for 2 categories, the distributions of $X^2_{[1:2]}$ and $X^2_{[2]}$ are chi-square distributions with degrees of freedom equal to $q(q+1)/2$ and $q(q-1)/2$ respectively.

Here **H** was presented as a matrix of constants. However, if we consider H_l as a random variable that takes on values h_{ls} with probability $\pi_s(\theta_a)$, where $\pi_s(\theta_a)$ is a probability under a Neyman smooth alternative hypothesis, then we can see that the test statistics is just a special case of the score statistic given by Rayner and Best. And further, Bartholomew (1987) showed that the joint probability function of the q-dimensional vector of binary variables can be uniquely expressed in terms of the $2^q - 1$ marginal probabilities form first to q-th order, which means

$$X^2_{[1:q]} = \chi^2_{PF}$$

However, in fact we only require fewer than $2^q - 1$ marginals to reproduce the Pearson-Fisher statistic for a composite null hypothesis since some residuals on the marginal are degenerate variables equal to zero due to linear dependencies among the rows of $\mathbf{H}_{[1:q]}$

for a composite null. Suppose there are g linear dependent rows. We can delete these g rows from $\mathbf{H}_{[1:q]}$ and denote the new matrix $\mathbf{H}_{[1:q;-g]}$. Then

$$X^2_{[1:q:-g]} = \chi^2_{PF}$$

II.3.4 Goodness of Fit Statistics

Assume c categories for each variable. Joreskog and Moustaki (2001) defined

$$GFfit^{(ij)} = n \sum_{ab} \frac{(p_{ab}^{(ij)} - \hat{\pi}_{ab}^{(ij)})^2}{\hat{\pi}_{ab}^{(ij)}}$$

$$i = 1, \cdots, q-1 \quad j = i+1, \cdots, q \quad a = 1, \cdots, c \quad b = 1, \cdots, c$$

Here the name GFfit stands for Goodness-of-fit.

Consider c kernel patterns $\boldsymbol{t}_g$, $g = 1, 2, \ldots, c$ that form, as columns, a c by c identity matrix, and consider the cq by T matrix $\boldsymbol{U}$ given by

$\boldsymbol{U}$

$$\begin{matrix} (\boldsymbol{t}_1 \otimes (\mathbf{1}_c \otimes \mathbf{1}_c \ldots \otimes \mathbf{1}_c) & \boldsymbol{t}_2 \otimes (\mathbf{1}_c \otimes \mathbf{1}_c \ldots \otimes \mathbf{1}_c) & \cdots & \boldsymbol{t}_c \otimes (\mathbf{1}_c \otimes \mathbf{1}_c \ldots \otimes \mathbf{1}_c) \\ = \mathbf{1}_c \otimes (\boldsymbol{t}_1 \otimes \mathbf{1}_c \ldots \otimes \mathbf{1}_c) & \mathbf{1}_c \otimes (\boldsymbol{t}_2 \otimes \mathbf{1}_c \ldots \otimes \mathbf{1}_c) & \cdots & \mathbf{1}_c \otimes (\boldsymbol{t}_c \otimes \mathbf{1}_c \ldots \otimes \mathbf{1}_c) \ldots \\ \mathbf{1}_c \otimes (\mathbf{1}_c \otimes \mathbf{1}_c \ldots \otimes \boldsymbol{t}_1) & \mathbf{1}_c \otimes (\mathbf{1}_c \otimes \mathbf{1}_c \ldots \otimes \boldsymbol{t}_2) & \ldots & \mathbf{1}_c \otimes (\mathbf{1}_c \otimes \mathbf{1}_c \ldots \otimes \boldsymbol{t}_c)) \end{matrix}$$

Note that linear dependencies exist among columns of $\boldsymbol{U}$; $\boldsymbol{V}$ consists of the linear independent columns of $\boldsymbol{U}$.Then a $c^2q(q-1)/2$ by T matrix $\boldsymbol{M}$ is given by

$$\boldsymbol{M}_{[2]}\begin{bmatrix} (u_1 \circ u_{c+1})' \\ (u_1 \circ u_{c+2})' \\ \vdots \\ (u_1 \circ u_{qc})' \\ (u_2 \circ u_{c+1})' \\ (u_2 \circ u_{c+2})' \\ \vdots \\ (u_2 \circ u_{qc})' \\ \vdots \\ (u_c \circ u_{c+1})' \\ (u_c \circ u_{c+2})' \\ \vdots \\ (u_c \circ u_{qc})' \\ \vdots \\ (u_{c+1} \circ u_{2c+1})' \\ \vdots \\ (u_{c+1} \circ u_{qc})' \\ \vdots \\ (u_{(q-2)c+1} \circ u_{(q-1)c+1})' \\ \vdots \\ (u_{(q-1)c} \circ u_{qc})' \end{bmatrix}$$

Linear dependencies exist among rows of $\boldsymbol{M}_{[2]}$; $\mathbf{H}_{[2]}$ consists of the linear independent rows of $\boldsymbol{M}_{[2]}$.

Then using $\boldsymbol{M}_{[2]}$, Cagnone and Mignani (2007) show that $GFfit^{(ij)}$ is a special case of $X^2_{[t:u]}$:

$$GFfit^{(ij)} = \boldsymbol{e}'(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{(ij)})^{+}\boldsymbol{e}$$

Where $\boldsymbol{A}^{+}$ indicates the Moore-Penrose generalized inverse of matrix $\boldsymbol{A}$, and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{(ij)} = n^{-1}\boldsymbol{\Omega}_{\boldsymbol{e}}$ with $\boldsymbol{\Omega}_{\boldsymbol{e}}$ evaluated at the MLE $\widehat{\boldsymbol{\theta}}$, and now

$$\boldsymbol{\Omega}_{\boldsymbol{e}} = \mathbf{M}_{[2]}^{(ij)}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')(\mathbf{M}_{[2]}^{(ij)})'$$

$\mathbf{M}_{[2]}^{(ij)}$ is a partition of the general matrix $\mathbf{M}_{[2]}$ such that

$$\mathbf{M}_{[2]}^{(ij)} = \begin{bmatrix} m'_{g+1} \\ m'_{g+2} \\ \vdots \\ m'_{g+c^2} \end{bmatrix}$$

Where $g = (\frac{(i-1)(2q-i)}{2} + j - i - 1)c^2$

If we apply the Pearson's χ^2 to the *2*-variable subset for variable i and variable j, the X_{ij}^2 statistic can be defined as follow:

$$X_{ij}^2 = X_{ij}^2(\hat{\theta}_{ij}) = n\boldsymbol{e}'(\widehat{\mathbf{D}}_{[2]}^{ij})^{-1}\boldsymbol{e}$$

$$\widehat{\mathbf{D}}_{[2]}^{ij} = \mathbf{H}_{[2]}^{(ij)} D(\hat{\pi}) \mathbf{H}_{[2]}^{(ij)'}$$

$$\boldsymbol{e} = \mathbf{H}_{[2]}^{(ij)}(\boldsymbol{p} - \widehat{\boldsymbol{\pi}}) \text{ is the matrix form of the marginal residuals}$$

However, under the null hypothesis, X_{ij}^2 does not distribute as chi-square since here the parameter θ_{ij} is estimated from full table. $GFfit^{(ij)}$ is actually the same as X_{ij}^2.

II.3.5 Orthogonal Components

Consider the $T - g - 1$ by 2^q matrix $\mathbf{H}^* = \boldsymbol{F}'\mathbf{H}_{[1:q;-g]}$, where g is the number of unknown model parameters to be estimated and $\mathbf{H}_{[1:q;-g]}$ is matrix $\mathbf{H}_{[1:q]}$ deleting g rows. $\mathbf{H}^*$ has full row rank. $\boldsymbol{F}$ is the upper triangular matrix such that $\boldsymbol{F}'\boldsymbol{\Omega_e}\boldsymbol{F} = \boldsymbol{I}$. $\boldsymbol{F} = (\boldsymbol{C}')^{-1}$, where $\boldsymbol{C}$ is the Cholesky factor of $\boldsymbol{\Omega_e}$. Premultiplication by $(\boldsymbol{C}')^{-1}$ orthonormalises the matrix $\mathbf{H}_{[1:q;-g]}$ relative to the matrix $D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$. Then

$$X_{PF}^2 = X_{[1:q;-g]}^2 = n\mathbf{r}'(\widehat{\mathbf{H}}^*)'\widehat{\mathbf{H}}^*\mathbf{r}$$

where $\widehat{\mathbf{H}}^* = \mathbf{H}^*(\widehat{\boldsymbol{\theta}})$, and $\mathbf{r} = (\widehat{\boldsymbol{p}} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}))$.

Define

$$\hat{\boldsymbol{\gamma}} = n^{\frac{1}{2}}\widehat{\boldsymbol{F}}'\mathbf{H}\mathbf{r} = n^{\frac{1}{2}}\widehat{\mathbf{H}}^*\mathbf{r}$$

where $\widehat{\boldsymbol{F}}$ is the matrix $\boldsymbol{F}$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. Then

$$X_{PF}^2 = \hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{j=T-g-1}} \hat{\gamma}_j^2$$

$\widehat{\mathbf{H}}^*\mathbf{r}$ has asymptotic covariance matrix $\boldsymbol{F'\Omega_e F = I_{T-g-1}}$. The elements $\hat{\gamma}_j^2$ are asymptotically independent chi-square random variables with $df = 1$ (Reiser, 2008). Components maybe obtained as sequential sum of squares. Redefine

$$z_s = \sqrt{n}\left(\pi_s(\widehat{\boldsymbol{\theta}})\right)^{-\frac{1}{2}}(\hat{p}_s - \pi_s(\widehat{\boldsymbol{\theta}})).$$

Perform the regression of $\boldsymbol{z}$ on the columns of $\mathbf{H}'$:

$$\boldsymbol{z = H'\beta}$$

Then,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{H}\widehat{\boldsymbol{W}}\mathbf{H}')^{-1}\mathbf{H}\widehat{\boldsymbol{W}}\mathbf{u}$$

where $\mathbf{u} = \sqrt{n}\mathbf{r}$, $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{D}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{1}{2}} = \widehat{\boldsymbol{D}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{1}{2}}$, and $\boldsymbol{D} = diag(\boldsymbol{\pi}(\boldsymbol{\theta}))$.

$\Sigma = \Sigma(\theta) = (\boldsymbol{I} - \boldsymbol{\pi}^{\frac{1}{2}}\left(\boldsymbol{\pi}^{\frac{1}{2}}\right)' - \boldsymbol{A}(\boldsymbol{A'A})^{-1}\boldsymbol{A'})$ is idempotent.

Let $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{1}{2}}\mathbf{H}'$. Then

$$\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{M}}'\widehat{\boldsymbol{M}})^{-1}\widehat{\boldsymbol{M}}'\boldsymbol{z}$$

$\hat{\gamma}_j^2, j = 1, T-g-1$ are the sequential SS from this regression. $\boldsymbol{\gamma = C'\beta}$ are the orthogonal coefficients.

Now define an orthogonal components version of $GFfit$:

$$GFfit_{\perp}^{(ij)} = \sum_{l=m+1}^{l=m+(c-1)^2} \hat{\gamma}_l^2$$

where $m = q + (i-1)(c-1)^2 + (j-2)(c-1)^2$.

Then

$$X^2_{[2]} = \sum_{i=1}^{i=q-1} \sum_{j=i+1}^{j=q} GFfit_{\perp}^{(ij)}$$

More generally,

$$X^2_{PF} = \sum_{l=1}^{l=q(c-1)} \hat{\gamma}_l^2 + \sum_{l=q(c-1)+1}^{l=\binom{q}{2}(c-1)^2} \hat{\gamma}_l^2 + \sum_{l=\binom{q}{2}(c-1)^2+1}^{l=\binom{q}{3}(c-1)^3} \hat{\gamma}_l^2 + \cdots + \hat{\gamma}^2_{T-g-1}$$

Then

$$X^2_{PF} = \sum_i GFfit_{\perp}^{(i)} + \sum_i \sum_j GFfit_{\perp}^{(ij)} + \sum_i \sum_j \sum_k GFfit_{\perp}^{(ijk)} + \cdots + GFfit_{\perp}^{(1,2,\cdots,q)}$$

because

$$X^2_{PF} = \hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}} = \sum_{l=1}^{l=T-g-1} \hat{\gamma}_l^2$$

The extended $GFfit_{\perp}^{(ij)}$ are independent chi-square statistics with $df = (c-1)^2$ because of the definition on orthogonal components. The original $GFfit^{(ij)}$ statistics are not necessarily independent and do not necessarily sum to $X^2_{[2]}$. $GFfit_{\perp}^{(ij)}$ statistics are order dependent since they are defined on orthogonal components. $GFfit_{\perp}^{(ij)}$ statistics are diagnostics for lack of fit. If $GFfit_{\perp}^{(23)}$ has a large value, it means the association between variable 2 and variable 3 cannot be explained by the current model.

Due to collinearity among the columns of H, the calculation of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-\mathbf{1}}$ is usually very inaccurate. Thus $X^2_{[2]}$ is very inaccurate numerically if we use $X^2_{[2]} = \boldsymbol{e}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-\mathbf{1}}\boldsymbol{e}$ to calculate it. However, calculating components by sequential SS as given in Section II.3.4 using the Sweep operator are very accurate numerically (Goodnight, 1978; SAS PROC REG).

II.4. Other Statistics

II.4.1 Joe-Maydeu Statistic

Since $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-\mathbf{1}}$ is usually very inaccurate, Maydeu-Olivares and Joe(2005) proposed an alternative quadratic form statistic. Suppose $\mathbf{H}_{[1:r]}$ is full rank with rank=s, Define

$$\boldsymbol{\Delta}_{\boldsymbol{r}} = \mathbf{H}_{[1:r]}\mathbf{G}$$

Then consider an $s\times(s-g)$ orthogonal complement to $\boldsymbol{\Delta}_{\boldsymbol{r}}$, say $\boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})}$, such that

$$\boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})\prime}\boldsymbol{\Delta}_{\boldsymbol{r}} = \mathbf{0}$$

Then let

$$C_r = C_r(\theta) = \boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})}[\boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})\prime}\mathbf{H}_{[1:r]}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\mathbf{H}_{[1:r]}{}'\boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})}]^{-\mathbf{1}}\boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})\prime}$$

Note that C_r is invariant to the choice of $\boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})}$. Since

$$\begin{aligned} C_r\boldsymbol{\Omega}_{\boldsymbol{e}}C_r &= C_r\mathbf{H}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-\mathbf{1}}\mathbf{G}')\mathbf{H}'C_r \\ &= C_r\mathbf{H}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\mathbf{H}'C_r - C_r\mathbf{H}(\mathbf{G}(\mathbf{A}'\mathbf{A})^{-\mathbf{1}}\mathbf{G}')\mathbf{H}'C_r \\ &= C_r\mathbf{H}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\mathbf{H}'C_r - 0 = C_r \end{aligned}$$

$\boldsymbol{\Omega}_{\boldsymbol{e}}$ is a generalized inverse of C_r.

Then the Joe-Maydeu statistic is defined as

$$M_r = M_r(\hat{\theta}) = n\boldsymbol{e}'\hat{C}_r\boldsymbol{e}$$

$$\hat{C}_r = C_r(\hat{\theta})$$

$\boldsymbol{e} = \mathbf{H}_{[1:r]}(\boldsymbol{p} - \boldsymbol{\pi})$ is the matrix form of the marginal residuals.

Under the null hypothesis, M_r is distributed asymptotically chi-square with $df = s - g$. The degrees of freedom are obtained using the fact that $\boldsymbol{\Delta}_{\boldsymbol{r}}^{(\boldsymbol{c})}$ is of full rank $s-g$ and hence C_r is also of rank $s-g$. Maydeu-Olivares and Joe have shown that M_q equals the Pearson-Fisher chi-square statistic when $\hat{\theta}$ is the maximum likelihood estimator.

However, when $\hat{\theta}$ is some other minimum variance asymptotically normal estimator, M_q and the Pearson-Fisher chi-square statistic are equivalent only asymptotically, with $M_q < X^2_{PF}$.

Maydeu-Olivares and Joe(2006) also proposed an $M_r^{(b)}$ statistic to assess the source of misfit when the overall M_r statistic suggest a model misfit. $M_r^{(b)}$ is based on each subset b of {1,…,n} with cardinality r. For a submodel for r-dimensional margins, with $C_r(b) = \prod_{i\in b} c_i$ cells depending on $g_r(b)$ parameters, $M_r^{(b)}$ has an asymptotic null chi-square distribution with $C_r(b) - g_r(b) - 1$ degrees of freedom, assuming that the submodel is identified, the estimator is consistent and asymptotically normal, and $C_r(b) - 1 > g_r(b)$. Let be $\theta_b$the subset of the parameter vector θ with dimension $g_r(b)$. Then the $M_r^{(b)}$ statistic is defined as

$$M_r^{(b)} = M_r^{(b)}\left(\hat{\theta}_b\right) = n\boldsymbol{e}'\hat{C}_{rb}\boldsymbol{e}$$

$$\hat{C}_{rb} = C_{rb}\left(\hat{\theta}_b\right) = \boldsymbol{\Delta}_{\boldsymbol{rb}}^{(\boldsymbol{c})}[\boldsymbol{\Delta}_{\boldsymbol{rb}}^{(\boldsymbol{c})\prime}\mathbf{H}_{[\mathrm{r}]}^{(b)}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\mathbf{H}_{[\mathrm{r}]}^{(b)\prime}\boldsymbol{\Delta}_{\boldsymbol{rb}}^{(\boldsymbol{c})}]^{-1}\boldsymbol{\Delta}_{\boldsymbol{rb}}^{(\boldsymbol{c})\prime}$$

$\boldsymbol{e} = \mathbf{H}_{[\mathrm{r}]}^{(b)}(\boldsymbol{p} - \boldsymbol{\pi})$ is the matrix form of the marginal residuals

$$\boldsymbol{\Delta}_{\boldsymbol{rb}} = \frac{\partial \mathbf{H}_{[\mathrm{r}]}^{(b)}\boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \theta_b}$$

$\boldsymbol{\Delta}_{\boldsymbol{rb}}^{(\boldsymbol{c})}$ *is* an $C_r(b) - 1 \times g_r(b)$ orthogonal complement to $\boldsymbol{\Delta}_{\boldsymbol{rb}}$

Given a necessary and sufficient condition that

$$\Sigma_{rb}\mathrm{C}_{rb}\Sigma_{rb}\mathrm{C}_{rb}\Sigma_{rb} = \Sigma_{rb}\mathrm{C}_{rb}\Sigma_{rb} \qquad for\ any\ \theta$$

where Σ_{rb} is the asymptotic covariance matrix of $\sqrt{n}\mathbf{H}_{[1:\mathrm{r}]}^{(b)}(\boldsymbol{p} - \boldsymbol{\pi})$, Maydeu-Olivares and Joe have shown that the $M_r^{(b)}$ has an asymptotic null chi-square distribution with $C_r(b) -$

$g_r(b) - 1$ degrees of freedom. The degrees of freedom are obtained by the fact that $\boldsymbol{\Delta}_{\boldsymbol{rb}}^{(\boldsymbol{c})}$ is of full column rank $C_r(b) - g_r(b) - 1$ and hence C_{rb} is also of rank $C_r(b) - g_r(b) - 1$.

II.4.2 $\bar{X}_{ij}^2$ and $\bar{\bar{X}}_{ij}^2$

As mentioned earlier X_{ij}^2 does not distributed chi-square. However, we can assume that the distribution of X_{ij}^2 can be approximated by a $b\chi_a^2$ distribution. The first and second asymptotic moments of X_{ij}^2 are

$$\hat{\mu}_1 = tr\left(\left(\widehat{\boldsymbol{D}}_{[2]}^{ij}\right)^{-1}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}\right), \;\; \hat{\mu}_2 = 2tr\left(\left(\widehat{\boldsymbol{D}}_{[2]}^{ij}\right)^{-1}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}\right)^2$$

Solving for the two unknown constants a and b, we obtain the mean and variance corrected $\bar{X}_{ij}^2$ statistic

$$\bar{X}_{ij}^2 = \frac{X_{ij}^2}{b} = \frac{2\hat{\mu}_1^2}{\hat{\mu}_2}X_{ij}^2$$

Which has an approximate reference chi-square distribution with degrees of freedom

$$a = \frac{2\hat{\mu}_1^2}{\hat{\mu}_2}$$

Alternatively, following Asparouhov and Muthen (2010), we can define a mean and variance corrected X_{ij}^2 which has $df_{ij} = c^2 - q_{ij} - 1$, where q_{ij} is the number of parameters in the bivariated probabilities. We can write the statistic $\bar{\bar{X}}_{ij}^2 = a^* + b^* X_{ij}^2$ where a^* and b^* are chosen so that the mean and variance of $\bar{\bar{X}}_{ij}^2$ are df_{ij} and $2df_{ij}$. Solving for a^* and b^*, we obtain

$$\bar{\bar{X}}_{ij}^2 = X_{ij}^2\sqrt{\frac{2df_{ij}}{\hat{\mu}_2}} + df_{ij} - \sqrt{\frac{2df_{ij}\hat{\mu}_1^2}{\hat{\mu}_2}}$$

II.4.3 Y Statistic

Bartholomew and Leung (2002) proposed the Y statistic based on only second order marginals. The statistic is defined as

$$\boldsymbol{Y} = \boldsymbol{e}'_{[2]}\widehat{\boldsymbol{D}}^{-1}_{[2]}\boldsymbol{e}_{[2]}$$

$$\boldsymbol{e}_{[2]} = \mathbf{H}_{[2]}(\widehat{\boldsymbol{p}} - \boldsymbol{\pi}(\boldsymbol{\theta}))$$

$$\boldsymbol{D}_{[2]} = \boldsymbol{n}^{-1}(\boldsymbol{diag}(\mathbf{H}_{[2]}\boldsymbol{\pi}(\boldsymbol{\theta}))(\boldsymbol{I} - \boldsymbol{diag}(\mathbf{H}_{[2]}\boldsymbol{\pi}(\boldsymbol{\theta}))))$$

Bartholomew and Leung gave a chi-square approximation with *c* degrees of freedom for the distribution of

$$\frac{Y - a}{b}$$

Where

$$b = \frac{\mu_3(Y)}{4\mu_2(Y)}$$

$$c = \frac{\mu_2(Y)}{2b^2}$$

$$a = \mu_1(Y) - bc$$

$$\mu_1, \mu_2 \; and \; \mu_3 \; are \; the \; asymptotic \; moments \; of \; Y$$

The Y statistic is simpler to compute than the $X^2_{[2]}$ since it only requires estimates for $\boldsymbol{\pi}$. However, this statistic does not perform well with the degrees of freedom given by Bartholomew and Leung. A modified version of this statistic, Y_2, was proposed by Cai, Maydeu, Coffman and Thissen(2006). Y_2 is based on both first and second order marginal and it is defined as

$$\boldsymbol{Y_2} = \boldsymbol{e}'_{[\mathbf{1:2}]}\widehat{\boldsymbol{D}}^{-1}_{[\mathbf{1:2}]}\boldsymbol{e}_{[\mathbf{1:2}]}$$

$$\boldsymbol{e}_{[\mathbf{1:2}]} = \mathbf{H}_{[\mathbf{1:2}]}(\widehat{\boldsymbol{p}} - \boldsymbol{\pi}(\boldsymbol{\theta}))$$

$$\boldsymbol{D_{[2]} = n^{-1}(diag(\mathbf{H_{[1:2]}}\pi(\theta))(I - diag(\mathbf{H_{[1:2]}}\pi(\theta))))}$$

A chi-square approximation with *c* degrees of freedom is given for the distribution of

$$\frac{Y_2 - a}{b}$$

Different from the Y statistic, now the computation of *a*, *b* and *c* require computation of $\boldsymbol{\Omega_e}$ evaluated at the maximum likelihood estimates $\hat{\pi}$ and $\hat{\theta}$. Thus the $\boldsymbol{Y_2}$ statistic has no computational advantage compared to $X^2_{[2]}$.

II.4.4 A "Reduced" Version of $\boldsymbol{X^2_{[t:u]}}$

Tollenaar and Mooijaart (2003) proposed a "reduced" version of $X^2_{[t:u]}$, X^2_{red}, defined as

$$X^2_{red} = n\boldsymbol{e}'(\mathbf{H_{[1:2]}}(D\left(\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\right) - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})')\mathbf{H_{[1:2]}}')^{-1}\boldsymbol{e}$$

where

$$\boldsymbol{e} = \mathbf{H_{[1:2]}}(\widehat{\boldsymbol{p}} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}))$$

The covariance matrix in X^2_{red} does not include the term $\mathbf{G(A'A)^{-1}G'}$, which may substantially reduce computations. The degrees of freedoms of X^2_{red} are different from those of $X^2_{[t:u]}$ because of the different covariance matrix. X^2_{red} has an asymptotic chi-square distribution with *m-g* degrees of freedom, where $m = 0.5q(q-1)$ and $g =$ number of parameters to be estimated. By substituting $\mathbf{H_{[1:r]}}$ in place of $\mathbf{H_{[1:2]}}$, X^2_{red} can be extended to include higher order marginal up to order *r*. The extended statistic is defined as

$$X^2_{red,r} = n\boldsymbol{e}'(\mathbf{H_{[1:r]}}(D\left(\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\right) - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})')\mathbf{H_{[1:r]}}')^{-1}\boldsymbol{e}$$

where

$$\boldsymbol{e} = \mathbf{H_{[1:r]}}(\widehat{\boldsymbol{p}} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}))$$

$X^2_{red,r}$ has an asymptotic chi-square distribution with *m-g* degrees of freedom, where $m =$ the rank of $\mathbf{H}_{[\mathbf{1}:\mathbf{r}]}$ and $g =$ number of parameters to be estimated. Note that $X^2_{red,r}$ is just the sum of $X^2_{(b)}$ statistics:

$$X^2_{red,r} = \sum_b X^2_{(b)}$$

II.5 Power of $GFfit^{(ij)}_{\perp}$

Suppose the true probability vector is π. Then we use a wrong model to fit the data and get the probability vector $\pi(\theta)$. Mitra (1958) shows that χ^2_{PF} has a limiting non-central chi-square distribution with non-centrality paremeter λ, where

$$\lambda = \delta' Diag[\pi(\theta)]^{-1}\delta$$

$$\pi = \pi(\theta) + \frac{\delta}{\sqrt{n}}$$

$GFfit^{(ij)}_{\perp}$ is calculated by decomposing χ^2_{PF} into orthogonal components γ^2 and $GFfit^{(ij)}_{\perp}$ is just the sum of several of these components. To calculate the power of $GFfit^{(ij)}_{\perp}$,we can apply the similar method introduced in section II.3.4.We can define the orthogonal components of λ. These orthogonal components may be used to calculate the power of $GFfit^{(ij)}_{\perp}$.

II.6 Generalized Linear Latent Variable Model

The generalized linear latent variable model (GLLVM) will be used for simulation and power calculation because the model is applied to a large number of variables and sparseness issues arise in large multidimensional tables. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_p)$ be the vector of p ordinal observed variables, each of them having c_i categories. Thus there are $\prod_{i=1}^{p} c_i$ cells, also called response patterns in the cross-classified table. The r-th response

pattern is indicated as $\boldsymbol{y_r} = (y_1 = a_1, y_2 = a_2, \dots, y_p = a_p)$, where a_i is the value of the i-th observed variable($a_i = 1, \dots, c_i \; and \; i = 1, \dots, p$).

Let $\boldsymbol{z} = (z_1, z_2, \dots, z_q)$ be the vector of q continuous latent variables. Then the probability of the r-th response pattern $\boldsymbol{y_r}$ is given by

$$\pi_r(\theta) = \int \pi_r(\boldsymbol{z})\, h(\boldsymbol{z})\, d\boldsymbol{z}\,,$$

where θ is a vector of parameters. $h(\boldsymbol{z})$ is the density function of z, and we assume every latent variable to be distributed standard normal independently. $\pi_r(z)$ is the conditional probability of y_r given z and it is a multinormial probability function

$$\pi_r(\boldsymbol{z}) = \prod_{i=1}^{p} \pi_{a_i}^{(i)}(z) = \prod_{i=1}^{p} (\tau_{a_i}^{(i)} - \tau_{a_{i-1}}^{(i)})$$

where $\tau_{a_i}^{(i)} = \pi_1^{(i)}(z) + \pi_2^{(i)}(z) + \cdots + \pi_{a_i}^{(i)}(z)$ is the probability of a response in category a_i or lower on the variable i and $\pi_{a_i}^{(i)}(z)$ is the probability of a response in category a_i on the variable i.

Logistic regression is used to model the interrelationship between $\tau_{a_i}^{(i)}$ and the latent variables.

$$log\left[\frac{\tau_s^{(i)}}{1 - \tau_s^{(i)}}\right] = \alpha_{i0}(s) - \sum_{j=1}^{q} \alpha_{ij} z_j\,, \;\; s = 1, \dots, c_{i-1}$$

$\alpha_{i0}(s)$ and α_{ij} are the parameters of the model. $\alpha_{i0}(s)$ is the intercept and α_{ij} is the j-th slope for variable i. The intercepts should satisfy the condition $\alpha_{i0}(1) \leq \alpha_{i0}(2) \leq \cdots \leq \alpha_{i0}(c_i)$.

We use the E-M algorithm to calculate the maximum likelihood estimator for the parameters in the model. The integrals are approximated through the Gauss-Hermite quadrature method (Cagnone & Mignani, 2007).

II.7 Completed Monte Carlo Simulations

I carried out several Monte Carlo simulations to compare the performance of Type I error and power of the different statistics discussed earlier. In particular, I first compared three global statistics, traditional Pearson chi-square statistic X_{PF}^2, second-order marginal statistic calculated by using the matrix inverse $X_{[2]inv}^2$ and second-order marginal statistic calculated by using sequential SS $X_{[2]ss}^2$.

II.7.1 Completed Type I Error Study

The empirical distribution under H_0 and the empirical Type I error rate were examined first because a statistic may not be useful if the Type I error rate is not close to the nominal level. If a statistic does not follow the hypothesized theoretical distribution due to a condition such as sparseness, then the empirical Type I error rate may not be close to the nominal level.

The design of this Type I error study is described as follows

- Model GLLVM with 1 latent factor
- Number of observed variables $p = 4, p = 5, p = 6$
- Number of categories for each variable $c = 3, c = 4$
- Number of samples 500
- Sample size $n = 500$

The intercepts range from -3 to 3 and are generated randomly. The factor loadings are the following: for $p = 4, \alpha_1 = (0.0, 0.1, 0.2, 0.6)'$; for $p = 5, \alpha_1 = (0.0, 3.0, 2.0, 1.0, 2.0)'$;

for $p = 6, \alpha_1 = (0.8, 0.7, 0.5, 0.3, 0.2, 0.1)'$. I got these parameters from a published paper. (Cagnone & Mignani, 2007).

When estimating the parameters, the procedure may not converge if any of the slopes are too large. We will omit the samples where any slope estimation is larger than four. This convergence problem does not happen in four and six variables cases. However, for five variables case, about 3.2% of the samples cannot be estimated. The software used is R.

In Table 1, the means and standard deviations of X^2_{PF}, $X^2_{[2]inv}$ and $X^2_{[2]ss}$ are reported. The Type I errors for each statistic are reported in Table 2. The tables show empirical Type I error for nominal $\alpha = 0.05$, using a chi-square distribution for each statistic.

TABLE 1: Mean and Standard Deviation of the Statistics $\boldsymbol{X^2_{PF}}$, $\boldsymbol{X^2_{[2]inv}}$ and $\boldsymbol{X^2_{[2]ss}}$

	Mean				Standard Deviation			
p	4	4	5	6	4	4	5	6
c	3	4	4	4	3	4	4	4
X^2_{PF}	68.14	240.93	999.51	4001.73	12.16	25.33	90.87	447.88
$X^2_{[2]inv}$	25.98	53.19	91.65	141.06	12.14	10.33	13.12	70.96
$X^2_{[2]ss}$	23.67	54.99	90.06	132.71	6.75	10.37	12.85	15.65
k	81	256	1024	4096	81	256	1024	4096
n/k	6.17	1.95	0.49	0.12	6.17	1.95	0.49	0.12

Note: k =Number of response patterns= c^p.

TABLE 2: Type I Error of the Statistics $\boldsymbol{X^2_{PF}}$, $\boldsymbol{X^2_{[2]inv}}$ and $\boldsymbol{X^2_{[2]ss}}$

	Type I error			
p	4	4	5	6
c	3	4	4	4
$\alpha = 0.05$				
X^2_{PF}	0.060	0.086	0.166	0.25
$X^2_{[2]inv}$	0.098	0.050	0.052	0.078
$X^2_{[2]ss}$	0.04	0.066	0.052	0.034

From these two tables, I can see that the Type I error of X^2_{PF} makes sense only for four variables three categories case because the sparseness problem is moderate here: $\frac{n}{k} = 6.17$ is still greater than 5. However, for all the other cases, the sparseness is quite severe, so that the Type I error of X^2_{PF} is very large. $X^2_{[2]inv}$ is very inaccurate numerically here, especially for the six variables case as it has a very large standard deviation compared to the $X^2_{[2]ss}$. The Type I error looks good for four variables four categories and five variables four categories cases. However, it is a bit large for four variables three categories and six variables four categories cases. $X^2_{[2]ss}$ is the best statistic in this simulation study. For four variables and five variables cases, its Type I errors are close to 0.05 and its standard deviations are not very large. However for six variables case its Type I error is a bit small. This may due to some of the 4 by 4 marginal tables are sparse. Table 3, Table 4, Table 5 and Table 6 shows the means of the orthogonal components $GFfit_{\perp}^{(ij)}$

TABLE 3: Mean of the $\boldsymbol{GFfit_{\perp}^{(ij)}}$, Four Variables

$GFfit_{\perp}^{(ij)}$	c = 3 Mean (df=4)	c = 4 Mean (df=9)
(43)	3.82	8.89
(42)	3.72	9.15
(41)	3.94	9.28
(32)	4.12	8.93
(31)	4.13	9.29
(21)	3.93	9.45

TABLE 4: Mean of the $\boldsymbol{GFfit}_{\perp}^{(ij)}$, Five Variables Four Categories

$GFfit_{\perp}^{(ij)}$	Mean (df=9)
(54)	9.01
(53)	9.26
(52)	9.01
(51)	9.31
(43)	8.62
(42)	8.94
(41)	9.15
(32)	8.99
(31)	9.12
(21)	8.64

TABLE 5: Mean of the $\boldsymbol{GFfit}_{\perp}^{(ij)}$, Six Variables Four Categories

$GFfit_{\perp}^{(ij)}$	Mean (df=9)
(65)	8.37
(64)	8.69
(63)	8.46
(62)	9.12
(61)	8.94
(54)	8.96
(53)	9.10
(52)	8.89
(51)	9.16
(43)	8.77
(42)	8.56
(41)	8.84
(32)	8.53
(31)	8.99
(21)	9.31

From these three tables, I can see that in each case, the means of every $GFfit_{\perp}^{(ij)}$ are close to each other. This is because within each case, $GFfit_{\perp}^{(ij)}$ are independent chi-squared statistics on $(c-1)^2$ degrees of freedom due to its definition. However, I did find that in the six variables four categories case, the empirical means of $GFfit_{\perp}^{(65)}$ and $GFfit_{\perp}^{(63)}$ are lower than what we expected. For a four categories case, the $GFfit_{\perp}^{(\mathrm{ij})}$

should distribute chi-squared with 9 degrees of freedom. But the empirical means of $GFfit_{\perp}^{(65)}$ and $GFfit_{\perp}^{(63)}$ are 8.37 and 8.46. This may due to the sparseness in the two-way subtable. The sum of $GFfit_{\perp}^{(ij)}$ equals to $X_{[2]ss}^2$ as shown in Section II.3.4. However, the original $GFfit^{(ij)}$ statistics are not necessarily independent and do not necessarily sum to $X_{[2]ss}^2$.

II.7.2 Completed Power Simulation Study

$X_{[2]}^2$ and $GFfit_{\perp}^{(ij)}$ may have higher power for certain alternative hypotheses because they represent a test that is "focused" on the second-order marginal. If lack of fit is present in second-order marginal, then $X_{[2]}^2$ and $GFfit_{\perp}^{(ij)}$ would have higher power than an omnibus statistic such as X_{PF}^2. But if lack of fit is present in higher-order marginal, then $X_{[2]}^2$ and $GFfit_{\perp}^{(ij)}$ may have lower power.

I used a four variables three categories dataset to study the power of X_{PF}^2, $X_{[2]inv}^2$ and $X_{[2]ss}^2$. This dataset has 500000 records for 1000 replications of samples of size 500. It is generated from the two-factor model. The correlation of these two factors is zero. The true intercepts are $\alpha_{10} = (-1.5, -0.6, 0.3, 1)'$ and $\alpha_{20} = (-1, -0.3, 0.6, 1.5)'$. The true slopes are $\alpha_1 = (1, 1, 1, 1)'$ and $\alpha_2 = (0, 0.1, 0.2, 0.6)'$. To evaluate the power, the one-factor model was used to fit the data. So the parameter vector is

$$\begin{bmatrix} \alpha_{101} \\ \alpha_{102} \\ \alpha_{103} \\ \alpha_{104} \\ \alpha_{201} \\ \alpha_{202} \\ \alpha_{203} \\ \alpha_{204} \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{bmatrix}$$

for the false one-factor model. As mentioned in Section 6.1, when estimating the parameters, the procedure may not converge if any of the slopes are too large. In this dataset, the convergence problem happens when any of the slopes are greater than 4.2. Besides the convergence problem, another problem happens when the start values of the intercepts differ too much. If α_{10} and α_{20} differ too much, the estimates of α_1 are tend to be greater than the estimates of α_2. These slope estimates will produce a negative probability, which does not make any sense. So this kind of sample will also be omitted. After omitting the samples with these two problems, there are 970 samples left. Our start values are $\alpha_{10} = (-0.9, -0.3, 0.3, 0.9)'$, $\alpha_{20} = (-0.6, 0, 0.6, 1.2)'$ and $\alpha_1 = (0.5, 0.5, 0.5, 0.5)'$. The Type I error is set to be 0.05. The mean, standard deviation and empirical power of these three statistics are shown in the following table. Empirical power is the number of samples that reject the null divided by 970.

TABLE 6: Mean, Standard Deviation and Power of $\boldsymbol{X^2_{PF}}$, $\boldsymbol{X^2_{[2]inv}}$ and $\boldsymbol{X^2_{[2]ss}}$

	Mean	Standard Deviation	Empirical Power
X^2_{PF}	84.53	25.47	0.268
$X^2_{[2]inv}$	72.98	100.28	0.800
$X^2_{[2]ss}$	39.74	10.23	0.591

From Table 6 we can see that the power of X^2_{PF} is 0.268, which is not large enough. $X^2_{[2]inv}$ has a very large power, which is 0.8. However, $X^2_{[2]inv}$ is very inaccurate numerically. This can be demonstrated by its large standard Deviation. $X^2_{[2]ss}$ is the best statistic here. Its power is 0.591 and it has a small standard deviation.The means of $GFfit^{(ij)}_{\perp}$ are reported in Table 7.

TABLE 7: Mean of the $\boldsymbol{GFfit^{(ij)}_{\perp}}$

$GFfit^{(ij)}_{\perp}$	Mean (df=4)
(43)	4.11
(42)	5.95
(41)	6.68
(32)	14.93
(31)	4.04
(21)	4.03

As mentioned earlier, $GFfit^{(ij)}_{\perp}$ should distribute chi-square on $(c-1)^2$ degrees of freedom independently if the model is correct. In this case, under the null hypothesis, $GFfit^{(ij)}_{\perp}$ should distribute chi-square on 4 degrees of freedom However, since we use a

one-factor model to fit the data generated by a two-factors model, we can see that the mean of $GFfit_{\perp}^{(32)}$ is 14.93, which is substantially higher than the other $GFfit_{\perp}^{(ij)}$. In the true model, variable 2 and 3 have factor loading 0.1 and 0.2 on factor 2.

CHAPTER 3

THEORETICAL AND EMPIRICAL STUDIES OF THE GFFIT STATISTIC

I studied three problems in my dissertation. Firstly, I studied the Type I error and power of $GFfit_{\perp}^{(ij)}$, both theoretical and empirical. Secondly, I improved the performance of $GFfit_{\perp}^{(ij)}$ when the two-way subtables are sparse. Thirdly, I applied the improvement on $GFfit_{\perp}^{(ij)}$ to $X_{[2]}^2$.

III.1 Type I Error and Power Study of $GFfit_{\perp}^{(ij)}$

I performed theoretical calculations to study asymptotic power and several Monte Carlo simulations to study the empirical Type I error and empirical power of $GFfit_{\perp}^{(ij)}$ and compared the performance of $GFfit_{\perp}^{(ij)}$ to that of $M_2^{(ij)}$, X_{ij}^2 and $\bar{\bar{X}}_{ij}^2$.

The empirical distribution under H_0 and the empirical Type I error rate were examined first because a statistic may not be useful if the Type I error rate is not close to the nominal level. If a statistic does not follow the hypothesized theoretical distribution due to a condition such as sparseness, then the empirical Type I error rate may not be close to the nominal level. For the Type I error study of $GFfit_{\perp}^{(ij)}$, $M_2^{(ij)}$, X_{ij}^2 and $\bar{\bar{X}}_{ij}^2$, sparseness in two-way subtables may affect the empirical distribution. Kolmogorov-Smirnov tests were applied to each statistic to test distribution against chi-square and performance at nominal α=0.01 and 0.05 were tabulated. As mentioned earlier, it is known that X_{ij}^2 is not distributed chi-square. The design of this Type I error study is as follows

- Model GLLVM with 1 latent factor
- Number of observed variables $p = 4, p = 5, p = 6$
- Number of categories for each variable $c = 3, c = 4$
- Number of samples 500
- Sample size $n = 500$ and 150

The intercepts are $\alpha_{0(i)} = (-1.5, 0.5)'$ for each variable for three categories case and $\alpha_{0(i)} = (-1.5, 0.5, 2.5)'$ for each variable for four categories. The factor loadings are the

following: for $p = 4, \alpha_1 = (0.0, 0.1, 0.2, 0.6)'$; for $p = 5, \alpha_1 = (0.0, 3.0, 2.0, 1.0, 2.0)'$; for $\mathrm{p} = 6, \alpha_1 = (0.8, 0.7, 0.5, 0.3, 0.2, 0.1)'$. These three parameter settings are the same as those introduced in the completed Type I error study of $X^2_{[2]}$. The true model was fitted to simulated data.

First, simulations with sample size 500 were conducted. Simulation results for Type I error are shown in the following tables. The tables show empirical Type I error rates for nominal $\alpha = 0.05$, using a chi-square distribution for each statistic. The error rates outside of the interval $0.05 \pm 1.96\sqrt{\frac{(0.95)(0.05)}{1000}} = (0.0365, 0.0635)$ were bolded.

Convergence rate for estimation of parameter for each case were also included in the table title.

TABLE 8: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables Three Categories, n=500, Convergence Rate=100%

	Type I error rate			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.038	0.05	**0.024**	0.036
(13)	**0.026**	0.058	**0.016**	0.040
(14)	0.05	0.056	**0.026**	0.040
(23)	0.054	0.058	**0.028**	0.046
(24)	0.04	0.066	**0.024**	0.042
(34)	0.05	0.044	**0.028**	0.046

TABLE 9: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables Four Categories, n=500, Convergence Rate=99%

	Type I error rate			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.048	0.051	0.038	0.055
(13)	0.061	0.058	0.056	**0.071**
(14)	0.040	0.034	**0.026**	0.048
(23)	0.038	0.040	0.036	0.048
(24)	0.068	0.057	0.042	0.069
(34)	**0.089**	**0.081**	0.060	**0.092**

TABLE 10: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Five Variables Four Categories, n=500, Convergence Rate=98%

	Type I error rate			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.053	0.053	0.0345	0.053
(13)	**0.071**	**0.071**	0.053	0.063
(14)	0.057	0.055	0.051	0.053
(15)	0.067	0.051	0.033	0.053
(23)	0.041	0.051	0.035	0.051
(24)	0.045	0.041	0.033	0.047
(25)	0.043	0.047	**0.027**	0.043
(34)	0.041	0.043	0.033	0.043
(35)	0.047	0.057	0.041	0.049
(45)	0.032	**0.029**	**0.022**	0.033

TABLE 11: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables Four Categories, n=500, Convergence Rate=100%

	Type I error rate			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.042	0.052	0.052	**0.072**
(13)	0.048	0.052	0.058	**0.070**
(14)	0.050	0.046	0.058	**0.070**
(15)	0.056	0.066	**0.070**	**0.072**
(16)	0.034	0.046	0.050	0.052
(23)	0.066	0.046	**0.072**	**0.084**
(24)	0.048	0.050	0.050	0.068
(25)	0.046	0.038	0.054	0.062
(26)	0.054	**0.080**	0.060	**0.070**
(34)	0.050	0.032	0.050	0.058
(35)	0.038	0.044	0.046	0.054
(36)	0.034	0.052	0.042	0.048
(45)	**0.030**	0.034	0.044	0.044
(46)	0.050	0.046	0.052	0.056
(56)	0.048	0.048	**0.078**	**0.080**

From these tables we can see that $GFfit_{\perp}^{(ij)}$, $M_2^{(ij)}$ and $\bar{\bar{X}}_{ij}^2$ have a good Type I error when sparseness is present but X_{ij}^2 does not since we mentioned in earlier chapters that X_{ij}^2 does not distribute Chi-square.

A Kolmogorov-Smirnov test has also been applied to each statistic. The p-values are shown in the following tables. I bolded the p-values less than 0.05.

TABLE 12: KS Test P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables Three Categories, n=500

	p-value			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	**<0.0001**	0.7279	**<0.0001**	**<0.0001**
(13)	**0.0006**	**0.0044**	**<0.0001**	**<0.0001**
(14)	0.5003	0.4846	**<0.0001**	**<0.0001**
(23)	0.8552	0.5093	**<0.0001**	**<0.0001**
(24)	0.7029	0.4102	**<0.0001**	**<0.0001**
(34)	0.6160	0.8930	**<0.0001**	**0.0002**

TABLE 13: KS Test P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables Four Categories, n=500

	p-value			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.1639	0.1036	**<0.0001**	0.6476
(13)	0.7483	0.8809	**<0.0001**	0.3310
(14)	0.6377	0.3924	**<0.0001**	0.8951
(23)	0.9001	0.8054	**<0.0001**	0.3214
(24)	0.7235	0.2745	**<0.0001**	**0.0226**
(34)	0.0181	0.4855	**<0.0001**	**0.0333**

TABLE 14: KS Test P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Five Variables Four Categories, n=500

	p-value			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.8100	0.8322	**<0.0001**	0.7614
(13)	0.0645	0.0334	**<0.0001**	**0.0173**
(14)	0.4235	0.3953	**<0.0001**	0.8063
(15)	0.1650	0.0384	**<0.0001**	**0.0094**
(23)	0.0762	0.8871	**<0.0001**	0.3721
(24)	0.9543	0.8644	**<0.0001**	0.9217
(25)	0.1810	0.2110	**<0.0001**	0.4059
(34)	0.6884	0.8972	**<0.0001**	0.6112
(35)	0.2433	0.9761	**<0.0001**	0.5721
(45)	0.2812	0.2106	**<0.0001**	0.0882

TABLE 15: KS Test P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables Four Categories, n=500

	p-value			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.0884	0.1357	**<0.0001**	0.6371
(13)	0.9253	0.9990	**<0.0001**	0.1950
(14)	0.0032	**0.0030**	**<0.0001**	0.0413
(15)	0.0431	0.1354	**<0.0001**	**0.0112**
(16)	0.0545	0.3613	**<0.0001**	0.8293
(23)	0.0842	0.6011	**<0.0001**	**0.0131**
(24)	0.5314	0.8327	**<0.0001**	0.2647
(25)	0.8498	0.7623	**<0.0001**	0.6642
(26)	0.4539	0.7237	**<0.0001**	0.5423
(34)	0.4711	0.4174	**<0.0001**	0.8828
(35)	0.0932	0.0842	**<0.0001**	0.2685
(36)	0.4498	0.7975	**<0.0001**	0.6673
(45)	0.2410	0.0924	**<0.0001**	0.6181
(46)	0.4129	0.3945	**<0.0001**	0.5516
(56)	0.6564	0.3897	**<0.0001**	0.2489

We can see that in all four scenarios, most of $GFfit_{\perp}^{(ij)}$ and $M_2^{(ij)}$ have a p-value greater than 0.05. None of the X_{ij}^2 has a p-value greater than 0.05, and theoretically it does not distribute Chi-squared. I noticed that for four variables four categories, five variables four

categories and six variables four categories cases, most $\bar{\bar{X}}_{ij}^2$ has a p-value greater than 0.05. However, for the four variables three categories case, none of the $\bar{\bar{X}}_{ij}^2$ has a p-value greater than 0.05. In this case, the degrees of freedom for $\bar{\bar{X}}_{ij}^2$ is 2. This result indicates that the $\bar{\bar{X}}_{ij}^2$ statistic does not approximate Chi-squared distribution well when the number of degrees of freedom is small.

Then I conducted another simulation with the same parameter settings but a smaller sample size of 150. With a smaller sample size, the contingency table is sparser than those in the earlier cases. With such a small sample size, some slope estimates tend to be very large, which indicate that the ML estimation algorithm for parameter estimates did not converge. An extremely large slope estimate will result in several estimated cumulative frequencies with the same value for different categories in one variable. In this case, we failed to compute the derivatives for the corresponding parameters. Without these derivatives, we cannot compute the statistics of interest. For example, for the four variables three categories case, using ltm package in R which produces MLE for parameters, 153 out of 500 samples have this problem. So I just discarded the four variables three categories case. The Simulation results for Type I error are shown in the following tables. The tables show empirical Type I error rates for nominal $\alpha = 0.05$, using a chi-squared distribution for each statistic.

TABLE 16: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables Four Categories, n=150, Convergence Rate=97.6%

	Type I error rate			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.041	0.037	0.027	0.043
(13)	0.055	0.051	0.041	0.057
(14)	0.055	0.047	0.035	0.041
(23)	0.039	0.041	**0.025**	**0.029**
(24)	0.047	0.033	0.031	0.047
(34)	0.047	0.047	0.029	0.053

TABLE 17: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Five variables Four Categories, n=150, Convergence Rate=98.8%

	Type I error rate			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.032	0.038	0.028	0.046
(13)	0.047	0.034	0.038	0.045
(14)	0.063	0.051	0.067	0.071
(15)	0.049	0.061	0.040	0.053
(23)	0.032	0.034	**0.022**	0.038
(24)	0.051	0.040	0.034	0.045
(25)	0.040	0.059	0.040	0.063
(34)	0.045	0.045	0.032	0.038
(35)	0.063	0.043	0.026	0.034
(45)	0.043	0.059	0.048	0.067

TABLE 18: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables Four Categories, n=150, Convergence Rate=99.8%

	Type I error rate			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.042	0.042	0.038	0.042
(13)	0.022	0.034	**0.014**	**0.028**
(14)	0.034	0.034	0.032	0.042
(15)	0.050	0.060	0.036	0.046
(16)	0.058	0.068	0.036	0.060
(23)	0.044	0.048	0.036	0.050
(24)	0.054	0.042	0.042	0.048
(25)	0.050	0.044	0.040	0.046
(26)	0.046	0.060	0.038	0.040
(34)	0.040	0.050	0.030	0.032
(35)	0.056	0.054	0.052	0.060
(36)	0.044	0.050	0.030	0.038
(45)	0.046	0.050	0.042	0.044
(46)	0.042	0.038	0.034	0.036
(56)	0.054	0.050	0.046	0.048

From these tables we can see that even with a smaller sample size and sparser contingency table, $GFfit_{\perp}^{(ij)}$, $M_2^{(ij)}$ and $\bar{\bar{X}}_{ij}^2$ have a good Type I error but X_{ij}^2 does not. A Kolmogorov-Smirnov test has also been applied to each statistic. The p-values are shown in the following tables.

TABLE 19: KS Test P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables Four Categories, n=150

	p-value			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.2258	0.0357	**<0.0001**	0.0303
(13)	0.6956	0.8877	**<0.0001**	0.8698
(14)	0.4740	0.8925	**<0.0001**	0.3517
(23)	0.0802	0.2740	**<0.0001**	0.5812
(24)	0.5394	0.4370	**<0.0001**	0.1872
(34)	0.5356	0.7980	**<0.0001**	0.3663

TABLE 20: KS Test P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Five Variables Four Categories, n=150

	p-value			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.4054	0.3125	**<0.0001**	0.3616
(13)	0.5599	0.2668	**<0.0001**	0.3065
(14)	0.2832	0.4409	**<0.0001**	0.1314
(15)	0.5160	0.3218	**<0.0001**	0.6426
(23)	0.2295	0.7410	**<0.0001**	0.6986
(24)	0.4844	0.8989	**<0.0001**	0.6365
(25)	0.5580	0.6263	**<0.0001**	0.7138
(34)	0.5057	0.8734	**<0.0001**	0.2494
(35)	0.0160	0.5289	**<0.0001**	0.3054
(45)	0.7394	0.7029	**<0.0001**	0.6587

TABLE 21: KS Test P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables Four Categories, n=150

	p-value			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.5718	0.7693	<0.0001	0.7006
(13)	0.1746	0.3033	<0.0001	**0.0014**
(14)	0.2841	0.6815	<0.0001	0.7957
(15)	0.7269	0.1877	<0.0001	0.1169
(16)	0.4218	0.6736	<0.0001	0.6599
(23)	0.4255	0.2892	<0.0001	0.1906
(24)	0.5222	0.4024	<0.0001	0.2571
(25)	0.9135	0.4619	<0.0001	0.9950
(26)	0.1849	0.6377	<0.0001	0.4906
(34)	0.2837	0.7848	<0.0001	0.6955
(35)	0.8376	0.8123	<0.0001	0.9370
(36)	0.6927	0.4109	<0.0001	0.5188
(45)	0.8323	0.6497	<0.0001	0.9616
(46)	0.1040	0.0277	<0.0001	0.0501
(56)	0.3942	0.2501	<0.0001	0.6346

We can see that in these three cases, most $GFfit_{\perp}^{(ij)}$, $M_2^{(ij)}$ and $\bar{\bar{X}}_{ij}^2$ have a p-value greater than 0.05.

According to these simulation studies, we can conclude that $GFfit_{\perp}^{(ij)}$, $M_2^{(ij)}$ and $\bar{\bar{X}}_{ij}^2$ still follow the hypothesized theoretical Chi-squared distribution even though there is a sparseness problem in the contingency table.

Besides these one-factor type I error rate study, I also studied the $GFfit_{\perp}^{(ij)}$ type I errors for two two-factor six variables four categories cases. For each case, pseudo data for 1000 samples were generated with sample size 500. The parameter settings are shown below: For the non-skewed case, $\alpha_{0(1)} = (-3, -2.5, -2, -1.8, -1.5, -0.8)'$, $\alpha_{0(2)} = (-1, -0.5, 0, 0.2, 0.5, 1.2)'$, $\alpha_{0(3)} = (1, 1.5, 2, 2.2, 2.5, 3.2)'$, $\alpha_1 = (1.6, 1.35, 1.25, 0.4, 0.5, 0.6)'$, $\alpha_2 = (0, 0, 0, 1, 1, 1)'$; for the skewed case, $\alpha_{0(1)} = (-3, -2.5, -2, -1.8, -1.5, -0.8)'$, $\alpha_{0(2)} = (-2.5, -2, -1.5, -1.3, -1, -0.3)'$, $\alpha_{0(3)} = (-2, -1.5, -1, -0.8, -0.5, 0.2)'$, $\alpha_1 = (1.6, 1.35, 1.25, 0.4, 0.5, 0.6)'$, $\alpha_2 = (0, 0, 0, 1, 1, 1)'$. I studied these two-factor cases because in the power study I generated the data from two-factor models and fitted the data with one-factor models. If the $GFfit_{\perp}^{(ij)}$ does not have a good type I error for two-factor cases, the power study would have no meaning. Simulation is available only for $GFfit_{\perp}^{(ij)}$ due to software. The type I error rates for these two-factor cases are listed below. The convergence rates for non-skewed case and skewed case are 99.5% and 99%, respectively.

TABLE 22: Type I Error Rates for Two-Factor Cases

	Type I error rates	
(ij)	Non-skewed	Skewed
(12)	0.0572	0.0626
(13)	0.0572	0.0535
(14)	0.05427	0.0454
(15)	0.03919	0.0434
(16)	0.0552	0.0474
(23)	0.0502	0.0636
(24)	0.0532	0.0545
(25)	0.0462	0.0484
(26)	0.0522	0.0515
(34)	0.0462	0.0454
(35)	0.0492	0.0636
(36)	0.0331	0.0434
(45)	0.0422	0.0515
(46)	0.0482	0.0474
(56)	0.0603	0.0383

From this table we can see that $GFfit_{\perp}^{(ij)}$ has good type I error for the two-factor cases. Besides 500 sample size, I also planned to do the simulation with sample size 150. However, with this small sample size, the convergence problem is presented so that I discarded these simulations. Beside the six variables case, I also planned to do a two-

factor four variables three categories type I error study. The parameters are $\alpha_{0(1)} = (-2,-2,-2,-2)', \alpha_{0(2)} = (2,2,2,2)', \alpha_1 = (0.0,1.0,1.0,0.0)', \alpha_2 = (2.0,0.1,0.2,2.0)'$. I wanted to investigate this case since I will use the same parameter setting to do a power study later. However, the convergence problem is present again for this case so that I have to discard the simulation.

Then I performed a power study for the lack-of-fit statistics examined in the type I error study. $X^2_{[2]}$, $GFfit^{(ij)}_{\perp}$, $M^{(ij)}_2$, X^2_{ij} and $\bar{\bar{X}}^2_{ij}$ may have higher power for certain alternative hypotheses because they represent a test that is "focused" on the second-order marginal. If lack of fit is present in second-order marginal, then these statistics may have higher power than an omnibus statistic such as X^2_{PF}. But if lack of fit is present in higher-order marginal, then these statistics may have lower power. For the power study, I calculated theoretical power of $GFfit^{(ij)}_{\perp}$ first, then compared to empirical power in the simulations. I tried several different cases in the simulation. In each case I examined the empirical power of $GFfit^{(ij)}_{\perp}$, $M^{(ij)}_2$, X^2_{ij} and $\bar{\bar{X}}^2_{ij}$.

First, I studied a four variables three categories case and six variables four categories case. Pseudo data for 1000 samples were generated from a confirmatory two-factor model with all parameters fixed and then fit with a one factor model. The parameters for the data generating models are the following: for four variables case, $\alpha_{0(1)} = (-2,-2,-2,-2)', \alpha_{0(2)} = (2,2,2,2)', \alpha_1 = (0.0,1.0,1.0,0.0)', \alpha_2 = (2.0,0.1,0.2,2.0)'$; for six variables case, $\alpha_{0(1)} = (-3,-2.5,-2,-1.8,-1.5,-0.8)', \alpha_{0(2)} = (-1,-0.5,0,0.2,0.5,1.2)', \alpha_{0(3)} = (1,1.5,2,2.2,2.5,3.2)', \alpha_1 = (1.6,1.35,1.25,0.4,0.5,0.6)', \alpha_2 = (0,0,0,1,1,1)'$, which are the same parameter used

for the Type I error study for a six-variable two-factor model. From the slopes, we can see that for the four variables case, variables 2 and 3 have high association with factor 1, variables 1 and 4 have high association with factor 2. For the six variables case, variables 1, 2 and 3 have high association with factor 1, variables 4, 5 and 6 have high association with factor 2. These parameter values are chosen so that the effect size for the goodness-of-fit test is large. The two latent variables were specified as uncorrelated, each with variance equal to 1.0. With a sparser dataset, the statistics tend to have a lower power. So for each case I used two different sample size, 150 and 500. The theoretical and empirical power for $GFfit_{\perp}^{(ij)}$ for the four-variable case are listed in the following table, and I bolded the relatively high power. The convergence rates for the sample size 500 case and sample size 150 case are 99.2% and 99.3%, respectively.

TABLE 23: Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$, Four Variables Case

	Power for $GFfit_{\perp}^{(ij)}$, four variables case			
	Sample size 500		Sample size 150	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	0.0548	0.0514	0.0514	0.0423
(13)	0.1341	0.1653	0.0726	0.0977
(14)	**0.4743**	**0.3891**	**0.1588**	**0.1319**
(23)	**0.6082**	**0.5806**	**0.2012**	**0.2134**
(24)	0.0501	0.0433	0.0500	0.0553
(34)	0.0502	0.0534	0.0501	0.0513

From this table we can see that with a smaller sample size, both theoretical and empirical power of $GFfit_{\perp}^{(ij)}$ tend to be smaller. The empirical power of $GFfit_{\perp}^{(ij)}$ is close to its theoretical power. For this four variables case, $GFfit_{\perp}^{(14)}$ and $GFfit_{\perp}^{(23)}$ have the two largest power, which shows that primarily the association between variables 2 and 3, and the association between variable 1 and 4, were not adequately explained by the one-factor model. For comparison, I also list the empirical power for $M_2^{(ij)}$, X_{ij}^2 and $\bar{\bar{X}}_{ij}^2$ in the following tables. I bolded relatively high power.

TABLE 24: Empirical Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables, n=500

	Empirical power			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.0514	0.0504	0.0353	0.0544
(13)	0.1653	0.0474	0.0272	0.0504
(14)	**0.3891**	0.0484	0.0221	0.0383
(23)	**0.5806**	0.1089	**0.8800**	**0.8740**
(24)	0.0433	0.0544	0.0272	0.0514
(34)	0.0534	0.0453	0.0292	0.0524

TABLE 25: Empirical Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Four Variables, n=150

	Empirical power			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	0.0423	0.0534	0.0342	0.0645
(13)	0.0977	0.0483	0.0322	0.0594
(14)	**0.1319**	0.0483	0.0292	0.0564
(23)	**0.2134**	0.0806	**0.3403**	**0.3353**
(24)	0.0553	0.0412	0.0362	0.0574
(34)	0.0513	0.0564	0.0392	0.0634

From these two tables we can see that $M_2^{(ij)}$ has very low power. Although X_{23}^2 has a large power, as I have demonstrated in the type I error study, it is not distributed Chi-squared. $\bar{\bar{X}}_{23}^2$ has a larger power than that of $GFfit_{\perp}^{(23)}$. However, the $\bar{\bar{X}}_{ij}^2$ statistic didn't detect the lack-of-fit in the associations between variable 1 and variable 4.

The theoretical and empirical powers for $GFfit_{\perp}^{(ij)}$ for six variables case are listed in the following table. The convergence rates for both sample size 500 case and sample size 150 case are 100%.

TABLE 26: Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$, Six Variables Case

	Power for $GFfit_{\perp}^{(ij)}$, six variables case			
	Sample size 500		Sample size 150	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	**0.2211**	**0.2280**	**0.0908**	**0.1160**
(13)	**0.2996**	**0.3240**	**0.1079**	**0.1300**
(14)	0.1000	0.0850	0.0634	0.0600
(15)	0.0964	0.0660	0.0625	0.0620
(16)	0.1188	0.1070	0.0679	0.0840
(23)	**0.9942**	**0.9630**	**0.5725**	**0.3850**
(24)	0.1196	0.0960	0.068	0.0630
(25)	0.1275	0.1240	0.0700	0.0630
(26)	**0.2306**	**0.1980**	**0.0929**	**0.0950**
(34)	**0.2374**	**0.2210**	**0.0944**	**0.1070**
(35)	0.1953	0.1820	0.0852	0.0900
(36)	0.1360	0.1070	0.0719	0.0570
(45)	0.0980	0.0670	0.0629	0.0650
(46)	0.1295	0.1180	0.0704	0.0650
(56)	0.1889	0.1540	0.0838	0.0850

Again, the theoretical power is close to the empirical power for $GFfit_{\perp}^{(ij)}$. For this six variables four categories case, $GFfit_{\perp}^{(23)}$ has the largest power, which means the

association between variable 2 and 3 cannot be explained by the one-factor model. When the sample size decreases, both theoretical power and empirical power decrease.

For comparison, I listed the empirical power for $M_2^{(ij)}$, X_{ij}^2 and $\bar{\bar{X}}_{ij}^2$ below.

TABLE 27: Empirical Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables, n=500

	Empirical power			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	**0.2280**	0.0460	0.0420	0.0660
(13)	**0.3240**	0.0490	0.0430	0.0630
(14)	0.0850	0.0610	0.0700	0.0940
(15)	0.0660	0.0380	0.0560	0.0710
(16)	0.1070	0.0610	0.0580	0.0820
(23)	**0.9630**	0.0460	0.0440	0.0570
(24)	0.0960	0.0610	0.0690	0.0900
(25)	0.1240	0.0530	0.0700	0.0860
(26)	0.1980	0.0570	0.0550	0.0770
(34)	0.2210	0.0510	0.0630	0.0830
(35)	0.1820	0.0490	0.0540	0.0740
(36)	0.1070	0.0420	0.0460	0.0630
(45)	0.0670	0.0510	**0.7210**	**0.7400**
(46)	0.1180	0.0470	**0.6610**	**0.6860**
(56)	0.1540	0.060	**0.6530**	**0.6740**

TABLE 28: Empirical Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables, n=150

	Empirical power			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	**0.1160**	0.0410	0.0460	0.0610
(13)	**0.1300**	0.0400	0.0410	0.0540
(14)	0.0600	0.0580	0.0460	0.0630
(15)	0.0620	0.0440	0.0420	0.0590
(16)	0.0840	0.0650	0.0540	0.0750
(23)	**0.3850**	0.0430	0.0260	0.0480
(24)	0.0630	0.0510	0.0490	0.0610
(25)	0.0630	0.0610	0.0440	0.0610
(26)	0.0950	0.0660	0.0470	0.0610
(34)	0.1070	0.0500	0.0460	0.0580
(35)	0.0900	0.0560	0.0460	0.0560
(36)	0.0570	0.0480	0.0360	0.0440
(45)	0.0650	0.0420	**0.190**	**0.2000**
(46)	0.0650	0.0420	**0.140**	**0.1490**
(56)	0.0850	0.0410	**0.161**	**0.1690**

From these two tables we can see that $M_2^{(ij)}$ has very low power. $\bar{\bar{X}}_{23}^2$ has a low power but $\bar{\bar{X}}_{45}^2$, $\bar{\bar{X}}_{46}^2$ and $\bar{\bar{X}}_{56}^2$ have relatively high power. This indicates that $GFfit_{\perp}^{(ij)}$ and $\bar{\bar{X}}_{ij}^2$ detect

the lack-of-fit in the associations between different pairs of variables. $GFfit_{\perp}^{(ij)}$ finds lack of fit in the first three variable pairs. $\bar{\bar{X}}_{ij}^{2}$ finds lack of fit in the last three variable pairs. Then I conducted two more power studies, still a four variables three categories case and a six variables four categories case to examine more sparseness conditions. First theoretical power was calculated, and then pseudo data for 1000 samples were generated from a confirmatory two-factor model with all parameters fixed and then fit with a one factor model. But the parameters are different from the earlier study: for four variables case, $\alpha_{0(1)} = (-1.5, -1, -0.6, -0.3)', \alpha_{0(2)} = (-1.0, -0.5, -0.1, 0.2)', \alpha_1 = (0.0, 1.0, 1.0, 0.0)', \alpha_2 = (2.0, 0.1, 0.2, 2.0)'$; for six variables case, $\alpha_{0(1)} = (-3, -2.5, -2, -1.8, -1.5, -0.8)', \alpha_{0(2)} = (-2.5, -2, -1.5, -1.3, -1, -0.3)', \alpha_{0(3)} = (-2, -1.5, -1, -0.8, -0.5, 0.2)', \alpha_1 = (1.6, 1.35, 1.25, 0.4, 0.5, 0.6)', \alpha_2 = (0, 0, 0, 1, 1, 1)'$. In the earlier settings, the intercepts range from -3 to 3.2, but in this setting almost all the intercepts are negative. The slopes in this setting are the same as in the earlier simulation. This will make the sample distribution in the two-way subtables more skewed and the statistics may have inaccurate Type I error level and/or lower power because of problems of higher sparseness in the subtables. Again, two sample sizes, 150 and 500 were used. For the four variables four categories case, the problem of failure to converge was encountered again using ltm package in R, and estimation did not converge in about half of the samples. For six-variable non-skewed case, all simulations converge for both 150 and 500 sample sizes. For six variables skewed case, 99% of simulations converge for 500 sample size and 92% of simulations converge for 150 sample size. So I

discard the four variables four categories case. The result for six variables four categories case is listed below.

TABLE 29: Power for $\boldsymbol{GFfit}_{\perp}^{(ij)}$, Six Variables Case

	Power for $GFfit_{\perp}^{(ij)}$, six variables case			
	Sample size 500		Sample size 150	
(*ij*)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	**0.3914**	**0.2790**	**0.1280**	**0.1654**
(13)	**0.5874**	**0.3984**	**0.1799**	**0.1513**
(14)	0.0648	0.0485	0.0542	0.0620
(15)	0.0616	0.0516	0.0533	0.0577
(16)	0.0706	0.0819	0.0558	0.0696
(23)	**0.8672**	**0.8413**	**0.3071**	**0.2546**
(24)	0.0614	0.0475	0.0533	0.0761
(25)	0.0565	0.0404	0.0519	0.0609
(26)	0.0576	0.0829	0.0522	0.0739
(34)	0.0752	0.0738	0.0571	0.0794
(35)	0.0671	0.0768	0.0549	0.0642
(36)	0.0546	0.0637	0.0514	0.0859
(45)	0.0592	0.0940	0.0527	0.0903
(46)	0.0683	0.0849	0.0552	0.0751
(56)	0.0697	0.0859	0.0556	0.0772

TABLE 30: Empirical Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables, n= 500

	Empirical power			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	**0.2790**	0.0758	0.1507	0.1799
(13)	**0.3984**	0.0758	0.1547	0.1789
(14)	0.0485	0.0374	0.0465	0.0596
(15)	0.0516	0.0384	0.0455	0.0596
(16)	0.0819	0.0566	0.0586	0.0758
(23)	**0.8413**	0.0647	0.1435	0.1658
(24)	0.0475	0.0404	0.0556	0.0707
(25)	0.0404	0.0434	0.0495	0.0647
(26)	0.0829	0.0475	0.0586	0.0697
(34)	0.0738	0.0394	0.0505	0.0697
(35)	0.0768	0.0505	0.0586	0.0788
(36)	0.0637	0.0485	0.0596	0.0738
(45)	0.0940	0.0889	**0.2912**	**0.3154**
(46)	0.0849	0.0495	**0.2436**	**0.2628**
(56)	0.0859	0.0394	**0.2386**	**0.2669**

TABLE 31: Empirical Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$, Six Variables, n= 150

	Empirical power			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	**0.1654**	0.0947	0.1360	0.1556
(13)	**0.1513**	0.0914	0.1175	0.1316
(14)	0.0620	0.0685	0.0663	0.0729
(15)	0.0577	0.0511	0.0577	0.0652
(16)	0.0696	0.0696	0.0642	0.0816
(23)	**0.2546**	0.0903	0.1055	0.1143
(24)	0.0761	0.0685	0.0772	0.0881
(25)	0.0609	0.0653	0.0631	0.0739
(26)	0.0739	0.0663	0.0609	0.0729
(34)	0.0794	0.0718	0.0707	0.0837
(35)	0.0642	0.0631	0.0653	0.0718
(36)	0.0859	0.0729	0.0707	0.0826
(45)	0.0903	0.0848	**0.1566**	**0.1697**
(46)	0.0751	0.0739	**0.1153**	**0.1218**
(56)	0.0772	0.0739	**0.1120**	**0.1273**

Again, from these tables, we can see that the theoretical power is close to the empirical power for $GFfit_{\perp}^{(ij)}$. $GFfit_{\perp}^{(ij)}$ and $\bar{\bar{X}}_{ij}^2$ detected the lack-of-fit in the associations between different pairs of variables.

Comparing the skewed case with the non-skewed case, we can see that with a more skewed dataset, $GFfit_{\perp}^{(ij)}$ tends to have lower power, both theoretically and empirically.

The empirical power is closer to the theoretical power in the non-skewed case than that in the skewed case. This difference may be due to more severe sparseness because of skewed table.

III.2 Improve $GFfit_{\perp}^{(ij)}$ by a Subset of Orthogonal Components

III.2.1 $GFfit_{\perp(\mathrm{t})}^{(ij)}$ Statistic

Although $GFfit_{\perp}^{(ij)}$ is a good remedy to the problem of sparseness because it is calculated from marginal two-way tables, sometimes even $GFfit_{\perp}^{(ij)}$ may have low power and inaccurate Type I error level due to severe sparseness in a two-way subtable when the number of categories is large and response variables have a skewed distribution. In that case, the distribution of $GFfit_{\perp}^{(ij)}$ may not be well approximated by the chi-square distribution even if the total sample size is large.

I modified $GFfit_{\perp}^{(ij)}$ for the sparse case by selecting a subset of orthogonal components chosen systematically to reduce the impact of sparseness to the extent possible. When computing $GFfit_{\perp}^{(ij)}$, we use $(c-1)^2$ orthogonal components that can produce the full table. Since including those orthogonal components corresponding to the cells with low frequencies is one reason for the poorer performance of $GFfit_{\perp}^{(ij)}$ in the sparse two-way subtable, one way to solve this problem is using a subset, less than $(c-1)^2$, of the orthogonal components corresponding to several cells with relatively large frequencies. In other words, instead of using all the $(c-1)^2$ components, we can drop those components that are likely to correspond to relatively small frequencies. I denote this statistic by $GFfit_{\perp(\mathrm{t})}^{(ij)}$, where t means computing the statistic with t cells, where $\mathrm{t} \leq$

$(c-1)^2$. Since $GFfit_{\perp(t)}^{(ij)}$ is the sum of *t* orthogonal components and each orthogonal component has a Chi-squared distribution with one degrees of freedom, $GFfit_{\perp(t)}^{(ij)}$ is distributed Chi-squared with *t* degrees of freedom. To use $GFfit_{\perp(t)}^{(ij)}$, we need to decide how many cells and which cells to choose to compute $GFfit_{\perp(t)}^{(ij)}$. Since including those cells with extremely low frequencies is the main reason that $GFfit_{\perp}^{(ij)}$ does not work well when the subtable is sparse, we should not choose too many cells. On the other hand, if we only choose 1 or 2 cells, we may decrease power of the test based on the component. So I investigated a moderate number of cells, say four or five cells, we seek to choose those cells with relatively large expected frequencies. I did many simulations and the best result is to choose the cells in the center of the table. The expected frequencies depend highly on the intercepts in the GLLVM model. Since we assume the latent variables are distributed normal in the model, if the intercepts are generally evenly distributed, then the cells in the center of the subtable will have large expected frequencies. For example, in the following tables I labeled the cells for a four categories case and a five categories case.

TABLE 32: Label of Cells for Four Categories Case.

Label		Category of variable i			
		1	2	3	4
Category of variable j	1	16	12	8	4
	2	15	11	7	3
	3	14	10	6	2
	4	13	9	5	1

TABLE 33: Label of Cells for Five Categories Case.

Label		Category of variable i				
		1	2	3	4	5
Category of variable j	1	25	20	15	10	5
	2	24	19	14	9	4
	3	23	18	13	8	3
	4	22	17	12	7	2
	5	21	16	11	6	1

For the four categories case, I will choose the four cells labeled 6, 7, 10, 11. For the five categories case, I will choose the five cells labeled 8, 12, 13, 14, 18. More generally, for a dataset with c categories in each variable, if c is even, I will choose four cells corresponding to the categories pair $(\frac{c}{2}, \frac{c}{2})$, $(\frac{c}{2}, \frac{c}{2}+1)$, $(\frac{c}{2}+1, \frac{c}{2})$ and $(\frac{c}{2}+1, \frac{c}{2}+1)$. If c is odd, I will choose five cells corresponding to the categories pair $(\frac{c+1}{2}, \frac{c+1}{2})$, $(\frac{c+1}{2}-1, \frac{c+1}{2})$, $(\frac{c+1}{2}, \frac{c+1}{2}-1)$, $(\frac{c+1}{2}+1, \frac{c+1}{2})$ and $(\frac{c+1}{2}, \frac{c+1}{2}+1)$. As the two-way table

becomes larger, more cells could be taken from the center of the table. For example, if the variable has six categories, then we can take 16 cells labeled "X" in the center of the two-way subtable as shown below.

TABLE 34: Cells to Choose to Compute $\boldsymbol{GFfit}_{\perp(\mathbf{t})}^{(\boldsymbol{ij})}$ for Six-Category Case

		Category of variable i					
		1	2	3	4	5	6
Category of variable j	1						
	2		X	X	X	X	
	3		X	X	X	X	
	4		X	X	X	X	
	5		X	X	X	X	
	6						

If the variable has seven categories, then we can take 13 cells labeled "X" in the center of the two-way subtable as shown below.

TABLE 35: Cells to Choose to Compute $\boldsymbol{GFfit}_{\perp(\mathbf{t})}^{(ij)}$ for Seven-Category Case

		Category of variable i						
		1	2	3	4	5	6	7
Category of variable j	1							
	2				X			
	3			X	X	X		
	4		X	X	X	X	X	
	5			X	X	X		
	6				X			
	7							

III.2.2 Type I Error Rate Study for $\boldsymbol{GFfit}_{\perp(\mathbf{t})}^{(ij)}$

To check the performance of the $GFfit_{\perp(\mathrm{t})}^{(ij)}$ statistic, I first conducted several Type I error studies. To demonstrate this sparseness problem, I conducted two Type I simulations for four variables, four categories. I generated 1000 pseudo datasets from a one factor model and fit it with a one factor model. The parameters for the data generating models are the following: $\alpha_{0(1)} = (-3.5, -3.5, -3.5, -3.5)', \alpha_{0(2)} = (0, 0, 0, 0)', \alpha_{0(3)} = (3.5, 3.5, 3.5, 3.5)', \alpha_1 = (1, 1, 1, 1)'$. The two sample sizes are 150 and 500. The average frequencies for each cells in the two-way subtables are listed below.

TABLE 36: Average Frequencies of Cells for Four Variables Four Categories Case, n=500

Average frequencies		Category of variable i			
		1	2	3	4
Category of variable j	1	2.13	13.02	6.61	0.414
	2	13.02	118.52	89.87	6.69
	3	6.67	89.46	118.23	13.05
	4	0.41	6.62	13.14	2.13

TABLE 37: Average Frequencies of Cells for Four Variables Four Categories Case, n= 150

Average frequencies		Category of variable i			
		1	2	3	4
Category of variable j	1	0.65	3.87	1.98	0.13
	2	3.86	35.80	26.95	1.98
	3	2.01	26.88	35.38	3.89
	4	0.12	1.96	3.90	0.63

Although the four cells in the middle have relatively large average frequencies, some of the other cells have very low frequencies, and with smaller sample size, the sparseness problem becomes more severe. Because of the sparseness in these cells, some $GFfit_{\perp}^{(ij)}$ statistics may have inaccurate empirical Type I error. The empirical Type I error rates of $GFfit_{\perp}^{(ij)}$ for these two cases when nominal $\alpha = 0.05$ are listed below.

TABLE 38: Type I Error Rates of $\boldsymbol{GFfit}_{\perp}^{(ij)}$ for Sparse Four Variables Four Categories Subtables.

	Type I error rate	
(ij)	Sample size 500	Sample size 150
(12)	**0.07**	**0.083**
(13)	**0.072**	**0.100**
(14)	0.057	**0.082**
(23)	0.047	**0.068**
(24)	0.054	**0.075**
(34)	**0.067**	**0.084**

Comparing to the interval $0.05 \pm 1.96\sqrt{\frac{(0.95)(0.05)}{1000}} = (0.0365, 0.0635)$, for the 500 sample size case, the empirical Type I error rates of $GFfit_{\perp}^{(12)}$, $GFfit_{\perp}^{(13)}$ and $GFfit_{\perp}^{(34)}$ are two high. With a sample size 150, the sparseness problem is so severe that all the empirical Type I error rates of $GFfit_{\perp}^{(ij)}$ are two high.

I applied the $GFfit_{\perp(\mathrm{t})}^{(ij)}$ to the four variables case with sparse two-way tables. Since the number of categories in this case is even, I chose four cells to compute $GFfit_{\perp(4)}^{(ij)}$. $GFfit_{\perp(4)}^{(ij)}$ is distributed asymptotically Chi-squared with 4 degrees of freedom. The empirical type I error rates for $GFfit_{\perp(4)}^{(ij)}$ are listed below.

TABLE 39: Type I Error Rates of $\boldsymbol{GFfit}_{\perp(\mathbf{4})}^{(\boldsymbol{ij})}$ for Sparse Four Variables Four Categories Subtables.

	Type I error rate	
(ij)	Sample size 500	Sample size 150
(12)	0.042	0.055
(13)	0.060	0.048
(14)	0.053	0.047
(23)	0.039	0.043
(24)	0.040	0.054
(34)	0.052	0.055

According to the result above, the empirical Type I error improved by using the four components corresponding to the four cells with the largest frequencies. All the empirical Type I error rates are within the interval $(0.0365, 0.0635)$, for both sample sizes.

A Kolmogorov-Smirnov test has also been applied to $GFfit_{\perp(4)}^{(ij)}$ to test its distribution against chi-square. The p-values are shown in the following table.

TABLE 40: KS Test P-values for $\boldsymbol{GFfit}_{\perp(\mathbf{4})}^{(\boldsymbol{ij})}$

	KS test p-values	
(ij)	Sample size 500	Sample size 150
(12)	0.384	0.176
(13)	0.316	**0.033**
(14)	0.354	0.652
(23)	0.378	0.566
(24)	0.449	0.668
(34)	0.931	0.411

For the 500 sample size case, all p-values are greater than 0.05. For the 150 sample size case, only for $GFfit_{\perp(4)}^{(13)}$, we reject the null hypothesis that it is distributed chi-square.

For further investigation, I conducted a type I error study for a five variables five categories case. I generated 1000 pseudo datasets from a one factor model and fit it with a one factor model. The parameters for the data generating models are the following: $\alpha_{0(1)} = (-3.5, -3.5, -3.5, -3.5, -3.5)', \alpha_{0(2)} = (-2.5, -2.5, -2.5, -2.5, -2.5)', \alpha_{0(2)} = (2.5,2.5,2.5,2.5,2.5)'$ $\alpha_{0(4)} = (3.5, 3.5, 3.5, 3.5)', \alpha_1 = (1, 1, 1, 1,1)'$. The sample size is 200. The expected frequencies for each cell in the two-way subtable are listed below.

TABLE 41: Average Frequencies of Cells for Five Variables Five Categories Case, n=200

Average frequencies		Category of variable i				
		1	2	3	4	5
Category of variable j	1	0.8548	1.6941	5.7310	0.4841	0.1642
	2	1.6884	3.7019	14.8257	1.4216	0.4945
	3	5.7027	14.8900	96.8331	14.8598	5.6709
	4	0.4831	1.4084	14.8639	3.7107	1.6623
	5	0.1684	0.4879	5.7018	1.6730	0.8237

From this subtable we can see that the cell with category 3 for both variable *i* and *j* has very large frequency. But all the other cells have relatively low frequencies. Since the number of categories is odd, when computing $GFfit_{\perp(t)}^{(ij)}$, I chose t=5. The empirical Type I error rates of both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(t)}^{(ij)}$ for these two cases when nominal $\alpha = 0.05$ are listed below.

TABLE 42: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ and $\boldsymbol{GFfit_{\perp(5)}^{(ij)}}$

	Type I error rates	
(ij)	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(5)}^{(ij)}$
(12)	**0.0752**	0.0542
(13)	**0.0742**	0.0621
(14)	**0.0682**	0.0471
(15)	**0.0682**	0.0542
(23)	**0.0742**	0.0611
(24)	**0.0812**	**0.0682**
(25)	**0.0862**	0.0632
(34)	**0.0822**	0.0421
(35)	**0.0662**	0.0502
(45)	**0.0672**	0.0451

From this table, we can see that all the empirical Type I error rates for $GFfit_{\perp}^{(ij)}$ are out of the interval $(0.0365,0.0635)$. But only one empirical Type I error rates for $GFfit_{\perp(5)}^{(ij)}$ are out of the interval $(0.0365,0.0635)$.

Choosing orthogonal components corresponding to cells with large frequencies can overcome the problem of sparseness in the two-way tables. Then, on the opposite, the $GFfit_{\perp(\mathrm{t})}^{(ij)}$ computed by choosing orthogonal components corresponding to cells with small frequencies will result in inaccurate type I error rates. For example, in the five variables five categories 200 sample size case, I also chose cells labeled 1, 3, 4, 11 and 16

to compute $GFfit_{\perp(5)}^{(ij)}$. The empirical Type I error rates of $GFfit_{\perp(5)}^{(ij)}$ for these chosen cells when nominal $\alpha = 0.05$ are listed below.

TABLE 43: Type I Error Rate for $\boldsymbol{GFfit_{\perp(5)}^{(ij)}}$ Choosing Cell 1, 3, 4, 11 and 16

$GFfit_{\perp(5)}^{(ij)}$	Type I error rate
(12)	**0.0652**
(13)	**0.0682**
(14)	**0.0682**
(15)	0.0621
(23)	**0.0692**
(24)	**0.0782**
(25)	**0.0852**
(34)	**0.0842**
(35)	0.0612
(45)	**0.0712**

Eight out of these ten Type I error rates are outside of the interval $(0.0365, 0.0635)$. From these simulations, we can see that when sparseness is present, using $GFfit_{\perp(t)}^{(ij)}$ may be a good remedy. However, even though the subtable is not sparse, $GFfit_{\perp(t)}^{(ij)}$ still distributed chi-squared distribution with df = t. To show this, I repeated the simulation study in chapter III.1 for the four variables four categories case with sample size 500. For

this case, the subtable is not sparse and I computed $GFfit_{\perp(4)}^{(ij)}$. The empirical type I error rates and KS test p-values are listed in the table below.

TABLE 44: Type I Error Rate for $\boldsymbol{GFfit_{\perp(4)}^{(ij)}}$

$GFfit_{\perp(4)}^{(ij)}$	Type I error rate	KS test p-value
(12)	**0.0283**	0.3822
(13)	0.0484	0.7685
(14)	0.0545	0.5125
(23)	0.0424	0.8549
(24)	0.0565	0.1681
(34)	0.0484	0.9759

With sample size 500, only $GFfit_{\perp(4)}^{(12)}$ has a type I error rate outside of the interval $0.05 \pm 1.96\sqrt{\frac{(0.95)(0.05)}{500}} = (0.0310, 0.0691)$. All the p-values are greater than 0.05.

III.2.3 Additional Type I Error Rate Study for $\boldsymbol{GFfit_{\perp(t)}^{(ij)}}$

In this section, I presented several additional Type I error rate simulation results for $GFfit_{\perp(t)}^{(ij)}$. The reason that I conducted these simulations is that the parameter settings used in these simulations are similar to the settings used in the power study for $GFfit_{\perp(t)}^{(ij)}$, which I will introduce later. If the $GFfit_{\perp(t)}^{(ij)}$ does not have a good Type I error rate in these simulations, the power study would have no meaning.

The first simulation is for a four-variable four-category case. 500 pseudo samples were generated from a one-factor model and fitted with a one-factor model. The parameters for the data generating model are the following: $\alpha_{0(1)} = (-1,-1,-1,-1)', \alpha_{0(2)} = (0.5,0.5,0.5,0.5)', \alpha_{0(3)} = (2,2,2,2)', \alpha_1 = (2.0, 1.1, 1.2, 2.0)'$. Two sample sizes are used, 500 and 150. The Type I error rates for $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$ are shown in the following table. The convergence rates for sample size 500 case and sample size 150 case are both 100%

TABLE 45: Type I Error Rates for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ and $\boldsymbol{GFfit_{\perp(4)}^{(ij)}}$, Four-Variable Four-Category

	Sample Size 500		Sample Size 150	
(ij)	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$
(12)	0.046	0.042	0.042	0.048
(13)	0.048	0.046	0.050	0.044
(14)	0.058	0.058	0.060	0.056
(23)	0.058	0.040	0.036	0.068
(24)	0.068	0.060	0.046	0.050
(34)	0.056	0.044	0.038	0.060

All these Type I error rates are within the interval $0.05 \pm 1.96\sqrt{\frac{(0.95)(0.05)}{500}} = (0.0310,0.0691)$. Thus both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$ works well for this case.

The second simulation is for a five-variable five-category case. 500 pseudo samples were generated from a one-factor model and fitted with a one-factor model. The sample size is 300. The parameters for the data generating model are the following: $\alpha_{0(1)} =$

$(-1.59, -2.30, -1.43, -3.02, -1.26)', \alpha_{0(2)} =$

$(-0.84, -0.38, -0.32, -1.50, -0.21)', \alpha_{0(3)} = (0.71,0.16,0.15,0.57,0.78)', \alpha_{0(4)} =$

$(1.48,1.80,1.66,2.13,1.65)', \alpha_1 = (2.3,2.5,1.9,2.1,2.3)'$. The Type I error rates for $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(5)}^{(ij)}$ are shown in the following table. The convergence rate is 100%.

TABLE 46: Type I Error Rates for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ and $\boldsymbol{GFfit_{\perp(5)}^{(ij)}}$, Five-Variable Five-Category

(ij)	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(5)}^{(ij)}$
(12)	0.036	0.056
(13)	0.050	0.058
(14)	0.048	0.048
(15)	0.060	0.060
(23)	0.040	0.040
(24)	0.048	0042
(25)	0.038	0.040
(34)	0.046	0.046
(35)	0.044	0.052
(45)	0.048	0.058

All these Type I error rates are within the interval $(0.0310,0.0691)$. Thus both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$ works well for this case.

The third simulation is for a four-variable six-category case. 500 pseudo samples were generated from a one-factor model and fitted with a one-factor model. The parameters for the data generating model are the following: $\alpha_{0(1)} = (-3.5, -3.5, -3.5, -3.5)', \alpha_{0(2)} =$

$(-3,-3,-3,-3)', \alpha_{0(3)} = (0,0,0,0)', \alpha_{0(4)} = (3,3,3,3)', \alpha_{0(5)} = (3.5,3.5,3.5,3.5)', \alpha_1 = (2.3, 2.5, 1.9, 2.1)'$. Two sample sizes are used, 1000 and 300. The Type I error rates for $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$ are shown in the following table. The convergence rates for sample size 1000 case and sample size 300 case are 98% and 98.6%, respectively.

TABLE 47: Type I Error Rates for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ and $\boldsymbol{GFfit_{\perp(4)}^{(ij)}}$, Four-Variable Six-Category

	Sample Size 1000		Sample Size 300	
(ij)	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$
(12)	**0.1020**	0.0592	**0.1156**	0.0568
(13)	**0.0857**	0.0388	**0.1115**	0.0507
(14)	**0.0816**	0.0429	**0.1338**	**0.0770**
(23)	**0.0694**	0.0469	**0.0872**	0.0486
(24)	**0.0837**	0.0531	**0.1075**	0.0568
(34)	**0.1040**	0.0673	**0.1014**	0.0649

We can see that for both 1000 sample size case and 500 sample size case, all the Type I error rates for $GFfit_{\perp}^{(ij)}$ are outside of the interval $(0.0310,0.0691)$ due to the sparseness in the two-way table. However all but one Type I error rates for $GFfit_{\perp(4)}^{(ij)}$ are within the interval $(0.0310,0.0691)$. This indicates that for this four-variable six-category case, $GFfit_{\perp(4)}^{(ij)}$ still distribute asymptotic chi-square but $GFfit_{\perp}^{(ij)}$ does not due to the sparseness in the two-way table.

The fourth simulation is for a five-variable five-category case. 500 pseudo samples were generated from a one-factor model and fitted with a one-factor model. The sample size is 300. The parameters for the data generating model are the following: $\alpha_{0(1)} = (-3,-3,-3,-3,-3)', \alpha_{0(2)} = (-2,-2,-2,-2,-2)', \alpha_{0(3)} = (2,2,2,2,2)', \alpha_{0(4)} = (3,3,3,3,3)', \alpha_1 = (2.5, 2.7, 1.9, 2.1, 2.3)'$.The Type I error rates for $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(5)}^{(ij)}$ are shown in the following table. The convergence rate is 100%.

TABLE 48: Type I Error Rates for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ and $\boldsymbol{GFfit_{\perp(5)}^{(ij)}}$, Five-Variable Five-Category

(ij)	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(5)}^{(ij)}$
(12)	0.058	0.050
(13)	**0.072**	0.048
(14)	0.060	0.048
(15)	**0.072**	0.050
(23)	0.052	0.052
(24)	**0.076**	**0.070**
(25)	0.062	0.038
(34)	0.064	0.052
(35)	0.054	0.038
(45)	0.062	0.064

For $GFfit_{\perp}^{(ij)}$, three out of then Type I error rates are outside of the interval $(0.0310,0.0691)$. But for $GFfit_{\perp(5)}^{(ij)}$, only one Type I error rates are outside of this

interval. This indicates that the $GFfit_{\perp}^{(ij)}$ does not work well due to the sparseness in the two-way subtable.

III.2.4 Power Study for $GFfit_{\perp(\mathbf{t})}^{(ij)}$

Besides the type I error study, I also conducted several power studies. When we use a wrong model to fit the data, both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(\mathrm{t})}^{(ij)}$ have a non-central chi-squared distribution. Using this property, we can compute the theoretical power of the $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(\mathrm{t})}^{(ij)}$. Although both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(\mathrm{t})}^{(ij)}$ distributed chi-squared, they may have different power. To show this, I conducted several power simulations, one four-variable four-category case and one five-variable five-category case. For the four-variable case, I used two sample sizes, 150 and 500. For the five-variable case, the sample size is 300. In both simulations, 500 pseudo samples were generated from a two-factor model and fitted with a one-factor model. The parameters for the data generating models are the following: for 4 variables case, $\alpha_{0(1)} = (-1,-1,-1,-1)', \alpha_{0(2)} = (0.5,0.5,0.5,0.5)', \alpha_{0(3)} = (2,2,2,2)', \alpha_1 = (0.0,1.0,1.0,0.0)', \alpha_2 = (2.0,0.1,0.2,2.0)'$; for 5 variables case, $\alpha_{0(1)} = (-1.59,-2.30,-1.43,-3.02,-1.26)', \alpha_{0(2)} = (-0.84,-0.38,-0.32,-1.50,-0.21)', \alpha_{0(3)} = (0.71,0.16,0.15,0.57,0.78)', \alpha_{0(3)} = (1.48,1.80,1.66,2.13,1.65)', \alpha_1 = (1.5,1.7,1.9,2.1,2.3)', \alpha_2 = (0.8,0.8,0,0,0)'$. These parameter settings are similar to the parameter settings used in the Type I error study shown in Sec III.2.2. The expected frequencies for each cell in the two-way subtables are listed below.

TABLE 49: Average Frequencies of Cells for Four-Variable Four-Category Case, n=500

Average frequencies		Category of variable i			
		1	2	3	4
Category of variable j	1	66.36	43.02	31.70	22.98
	2	42.97	35.09	28.93	23.12
	3	31.68	28.94	26.50	23.34
	4	22.99	23.15	23.38	25.77

TABLE 50: Average Frequencies of Cells for Four-Variable Four-Category Case, n=150

Average frequencies		Category of variable i			
		1	2	3	4
Category of variable j	1	19.91	12.90	9.51	6.89
	2	12.89	10.52	8.68	6.93
	3	9.50	8.68	7.95	7.00
	4	6.89	6.94	7.01	7.73

TABLE 51: Average Frequencies of Cells for Five-Variable Five-Category Case, n=150

Average frequencies		Category of variable i				
		1	2	3	4	5
Category of variable j	1	32.99	16.18	11.79	8.05	5.63
	2	13.76	10.04	9.66	7.84	6.91
	3	10.14	9.66	9.30	10.36	10.09
	4	6.79	7.74	9.76	11.43	15.52
	5	4.83	6.83	10.56	16.20	37.85

We can see that the four-variable 500 sample size case is not sparse, but the four-variable 150 sample size case and the five-variable case may be a little bit sparse but the sparseness is not severe. The theoretical and empirical power for both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(\mathrm{t})}^{(ij)}$ are listed below.

TABLE 52: Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ and $\boldsymbol{GFfit_{\perp(\mathbf{t})}^{(ij)}}$, Four-Variable Four-Category, n=500

	$GFfit_{\perp}^{(ij)}$		$GFfit_{\perp(4)}^{(ij)}$	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	0.0555	0.0681	0.0546	0.0641
(13)	0.0619	0.0701	0.0635	0.0841
(14)	**0.4513**	**0.3507**	**0.3615**	**0.3146**
(23)	**0.8271**	**0.8096**	**0.5611**	**0.5751**
(24)	0.0501	0.0501	0.0500	0.0400
(34)	0.0504	0.0621	0.0500	0.0600

TABLE 53: Power for $\boldsymbol{GFfit}_{\perp}^{(\boldsymbol{ij})}$ and $\boldsymbol{GFfit}_{\perp(\mathbf{t})}^{(\boldsymbol{ij})}$, Four-Variable Four-Category, n=150

	$GFfit_{\perp}^{(ij)}$		$GFfit_{\perp(4)}^{(ij)}$	
(*ij*)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	0.0516	0.0668	0.0513	0.0587
(13)	0.0534	0.0566	0.0539	0.0506
(14)	**0.1430**	**0.1356**	**0.1284**	**0.1417**
(23)	**0.2794**	**0.2692**	**0.1852**	**0.1741**
(24)	0.0500	0.0506	0.0500	0.0465
(34)	0.0500	0.0769	0.0500	0.0627

TABLE 54: Powers for $\boldsymbol{GFfit}_{\perp}^{(ij)}$ and $\boldsymbol{GFfit}_{\perp(\mathbf{t})}^{(ij)}$, Five-Variable Five-Category, n=300

	$GFfit_{\perp}^{(ij)}$		$GFfit_{\perp(5)}^{(ij)}$	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	**0.1638**	**0.1470**	**0.1209**	**0.1020**
(13)	0.0682	0.0570	0.0567	0.0440
(14)	0.0771	0.0630	0.0641	0.0580
(15)	0.0581	0.0480	0.0571	0.0480
(23)	0.0545	0.0620	0.0540	0.0450
(24)	0.0701	0.0450	0.0575	0.0510
(25)	0.0526	0.0620	0.0513	0.0410
(34)	0.0553	0.0500	0.0536	0.0570
(35)	0.0514	0.0580	0.0516	0.0400
(45)	0.0570	0.0490	0.0531	0.0480

From these tables, we can make three conclusions. First, when the subtable is not sparse, both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(\mathrm{t})}^{(ij)}$ have a non-central chi-squared distribution. This is demonstrated by the fact that the theoretical power of both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(\mathrm{t})}^{(ij)}$ are close to their empirical power. Second, when there is no sparse problem in the subtable, generally the power of $GFfit_{\perp(\mathrm{t})}^{(ij)}$ is lower than that of $GFfit_{\perp}^{(ij)}$. For example, in the four variables four categories 500 sample size case, $GFfit_{\perp}^{(23)}$ has a theoretical power of 0.8271, which is higher than the theoretical power of $GFfit_{\perp(4)}^{(23)}$, 0.5611. Third, even though there is no sparse problem in the subtable, the theoretical powers of $GFfit_{\perp}^{(ij)}$ and

$GFfit_{\perp(\mathrm{t})}^{(ij)}$ will decrease when the sample size decreases. For example, in the four variables four categories 500 sample size case, $GFfit_{\perp(4)}^{(23)}$ has a theoretical power of 0.5611. However, when sample size decreases to 150, $GFfit_{\perp}^{(23)}$ has a theoretical power of 0.1852.

However, when we have a sparse subtable, the power of $GFfit_{\perp}^{(ij)}$ might be lower than that of $GFfit_{\perp(\mathrm{t})}^{(ij)}$. To show this, 500 pseudo samples were generated from a two-factor model and fitted with a one-factor model. There are four variables and six categories in the dataset. The sample size is 1000. The parameters for the data generating models are the following: $\alpha_{0(1)} = (-3.5, -3.5, -3.5, -3.5)', \alpha_{0(2)} = (-3, -3, -3, -3)', \alpha_{0(3)} = (0,0,0,0)', \alpha_{0(4)} = (3,3,3,3)', \alpha_{0(5)} = (3.5,3.5,3.5,3.5)', \alpha_1 = (1.5, 1.7, 1.9, 2.1)', \alpha_2 = (0.8, 0.8, 0, 0)'$. This parameter setting is similar the parameter settings used in the Type I error study shown in Sec III.2.2. The expected frequencies for each cell in the two-way subtable are listed below.

TABLE 55: Average Frequencies of Cells for Four-Variable Six-Categories Case, n=1000

Average frequencies		Category of variable i					
		1	2	3	4	5	6
Category of variable j	1	26.21	7.48	44.91	14.49	0.55	0.95
	2	6.39	2.35	17.89	7.27	0.30	0.52
	3	36.05	16.07	178.43	121.32	6.56	12.19
	4	12.19	6.56	121.32	178.43	16.07	36.05
	5	0.52	0.30	7.27	17.89	2.35	6.39
	6	0.95	0.55	14.49	44.91	7.48	26.21

We can see that even though we have a large sample size of 1000, with the given parameter setting, there are still many cells in the subtable have a frequency less than 1. Thus there is a problem of sparseness in the subtable. The four cells in the center of the table have relatively large frequencies. In Sec III.2.2, I have shown that for this case $GFfit_{\perp}^{(ij)}$ does not work well but $GFfit_{\perp(4)}^{(ij)}$ works better. The theoretical and empirical powers for both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$ are listed below.

TABLE 56: Power for $\boldsymbol{GFfit}_{\perp}^{(\boldsymbol{ij})}$ and $\boldsymbol{GFfit}_{\perp(\mathbf{t})}^{(\boldsymbol{ij})}$, Four-Variable Six-Category, n=1000

	$GFfit_{\perp}^{(ij)}$		$GFfit_{\perp(4)}^{(ij)}$	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	0.3319	0.2953	0.5404	0.5213
(13)	0.0571	0.1140	0.0585	0.0835
(14)	0.0596	0.1283	0.0680	0.1059
(23)	0.0564	0.0875	0.0568	0.0733
(24)	0.0534	0.0733	0.0521	0.0794
(34)	0.0522	0.0631	0.0514	0.0753

We can see that when sparseness is presented in this subtable, both theoretical power and empirical power of $GFfit_{\perp}^{(12)}$ are lower than those of $GFfit_{\perp(4)}^{(12)}$. I already showed that in Sec III.2.2, $GFfit_{\perp}^{(12)}$ does not work well for this case due to the sparseness in the two-way table. This can also be demonstrated by the fact that the empirical power of $GFfit_{\perp}^{(13)}$ and $GFfit_{\perp}^{(14)}$ are about twice as their theoretical power.

If we decrease the sample size to 300, The theoretical and empirical powers for both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$ are listed below.

TABLE 57: Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ and $\boldsymbol{GFfit_{\perp(4)}^{(ij)}}$, Four-Variable Six-Category, n=300

	$GFfit_{\perp}^{(ij)}$		$GFfit_{\perp(4)}^{(ij)}$	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	0.1094	0.1000	0.1785	0.1700
(13)	0.0521	0.106	0.0526	0.0520
(14)	0.0528	0.080	0.0553	0.0480
(23)	0.0519	0.086	0.0521	0.0440
(24)	0.0510	0.102	0.0506	0.0480
(34)	0.0507	0.076	0.0504	0.0500

From these results, we can see that both theoretical and empirical powers of $GFfit_{\perp(4)}^{(12)}$ are still higher than those of $GFfit_{\perp}^{(12)}$. With a smaller sample size, the theoretical power of both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(\mathrm{t})}^{(ij)}$ are lower than those with a large sample size.

Then I did another simulation with a five variables five categories case. Again 500 pseudo samples were generated from a two-factor model and fitted with a one-factor model. The sample size is 150. The parameters for the data generating models are the following: $\alpha_{0(1)} = (-3,-3,-3,-3,-3)', \alpha_{0(2)} = (-2,-2,-2,-2,-2)', \alpha_{0(3)} = (2,2,2,2,2)', \alpha_{0(4)} = (3,3,3,3,3)', \alpha_1 = (1.5, 1.7, 1.9, 2.1, 2.3)', \alpha_2 = (1.0, 1.0, 0, 0, 0)'$, which is similar to the parameter setting used in the third simulation shown in Sec III.2.2. The expected frequencies for each cell in the two-way subtable are listed below

TABLE 58: Average Frequencies of Cells for Five-Variable Five-Category Case, n=150

Average frequencies		Category of variable i				
		1	2	3	4	5
Category of variable j	1	14.41	7.10	18.16	1.01	0.75
	2	5.93	4.38	15.97	1.19	0.93
	3	14.86	14.85	100.80	14.85	14.86
	4	0.93	1.19	15.97	4.38	5.93
	5	0.75	1.01	18.16	7.10	14.41

From this table we can see that actually only one cell has large expected frequencies and all the other cells have relatively low expected frequencies. The theoretical and empirical powers for both $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(5)}^{(ij)}$ are listed below. The convergence rate is 100%.

TABLE 59: Powers for $\boldsymbol{GFfit}_{\perp}^{(ij)}$ and $\boldsymbol{GFfit}_{\perp(\mathbf{t})}^{(ij)}$, Five-Variable Five-Category n=300

	$GFfit_{\perp}^{(ij)}$		$GFfit_{\perp(4)}^{(ij)}$	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	0.3286	0.226	0.3563	0.308
(13)	0.0546	0.068	0.0530	0.070
(14)	0.0541	0.052	0.534	0.040
(15)	0.0562	0.064	0.0574	0.052
(23)	0.0532	0.078	0.0517	0.058
(24)	0.0529	0.058	0.0516	0.050
(25)	0.0529	0.072	0.0514	0.048
(34)	0.0507	0.072	0.0510	0.054
(35)	0.0509	0.054	0.0510	0.068
(45)	0.0509	0.050	0.0510	0.052

From this table we can see that the theoretical powers of $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(5)}^{(ij)}$ are almost the same in this case, which means theoretically $GFfit_{\perp(5)}^{(ij)}$ actually did not improve the original $GFfit_{\perp}^{(ij)}$ much. But the empirical power for $GFfit_{\perp(5)}^{(12)}$ is higher than the empirical power for $GFfit_{\perp}^{(12)}$. This may be due to the poor performance of $GFfit_{\perp}^{(ij)}$ for this case because of the sparseness in the two-way subtable as demonstrated in Sec III.2.2. All the other $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(5)}^{(ij)}$ have very low power, both empirically and theoretically.

Then I applied $GFfit_{\perp(\mathrm{t})}^{(ij)}$ to the six variables four categories power study in Sec III.1.2. I chose four cells in each two-way subtable to compute $GFfit_{\perp(4)}^{(ij)}$ and compared the power for $GFfit_{\perp(\mathrm{t})}^{(ij)}$ with the power for $GFfit_{\perp}^{(ij)}$ in each case. The results for both non-skewed case and skewed case are shown in the following tables.

TABLE 60: Empirical Power for $\boldsymbol{GFfit}_{\perp}^{(\boldsymbol{ij})}$ and $\boldsymbol{GFfit}_{\perp(\mathbf{4})}^{(\boldsymbol{ij})}$, Non-Skewed Case.

	Powers for $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$			
(ij)	Sample size 500		Sample size 150	
	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$
(12)	**0.228**	**0.185**	**0.116**	**0.097**
(13)	**0.324**	**0.240**	**0.130**	**0.098**
(14)	0.085	0.079	0.060	0.050
(15)	0.066	0.069	0.062	0.060
(16)	0.107	0.101	0.084	0.071
(23)	**0.963**	**0.917**	**0.385**	**0.337**
(24)	0.096	0.083	0.063	0.061
(25)	0.124	0.090	0.063	0.057
(26)	0.198	0.197	0.095	0.095
(34)	0.221	0.205	0.107	0.100
(35)	0.182	0.170	0.090	0.086
(36)	0.107	0.122	0.057	0.077
(45)	0.067	0.100	0.065	0.075
(46)	0.118	0.134	0.065	0.075
(56)	0.154	0.196	0.085	0.099

TABLE 61: Empirical Power for $\boldsymbol{GFfit}_{\perp}^{(ij)}$ and $\boldsymbol{GFfit}_{\perp(4)}^{(ij)}$, Skewed Case.

	Powers for $GFfit_{\perp}^{(ij)}$ and $GFfit_{\perp(4)}^{(ij)}$			
(ij)	Sample size 500		Sample size 150	
	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$
(12)	**0.2790**	**0.0788**	**0.1654**	**0.1001**
(13)	**0.3984**	**0.0930**	**0.1513**	**0.0837**
(14)	0.0485	0.0293	0.0620	0.0663
(15)	0.0516	0.0415	0.0577	0.0500
(16)	0.0819	0.0424	0.0696	0.0642
(23)	**0.8413**	**0.1344**	**0.2546**	**0.1109**
(24)	0.0475	0.0434	0.0761	0.0565
(25)	0.0404	0.0414	0.0609	0.0598
(26)	0.0829	0.0485	0.0739	0.0598
(34)	0.0738	0.0374	0.0794	0.0729
(35)	0.0768	0.0576	0.0642	0.0783
(36)	0.0637	0.0374	0.0859	0.0685
(45)	0.0940	0.0889	0.0903	0.0946
(46)	0.0849	0.0728	0.0751	0.0729
(56)	0.0859	0.0616	0.0772	0.0805

From these two tables we can see that both $GFfit_{\perp(4)}^{(ij)}$ and $GFfit_{\perp}^{(ij)}$ have the largest power for variables pair 2 and 3. For the non-skewed case, the empirical power of

$GFfit_{\perp(4)}^{(23)}$ is just a little bit lower than the empirical power of $GFfit_{\perp}^{(23)}$. However, for the skewed case with sample size 500, the empirical power of $GFfit_{\perp(4)}^{(23)}$ is much lower than the empirical power of $GFfit_{\perp}^{(23)}$. To further investigate the reason for such a difference, I listed the expected frequencies of cells in the two-way table for the skewed case with sample size 500 below.

TABLE 62: Expected Frequencies of Cells for Skewed Case, Sample Size 500

Average frequencies		Category of variable i			
		1	2	3	4
Category of variable j	1	26.57	9.01	10.29	76.78
	2	6.75	2.59	3.10	27.24
	3	6.89	2.74	3.34	31.61
	4	31.11	13.56	17.36	230.98

We can see that due to the high skewness in the two-way subtable, the four cells in the center which we choose to compute $GFfit_{\perp(4)}^{(ij)}$ actually have lowest expected frequencies among all these cells. Thus, the $GFfit_{\perp(4)}^{(ij)}$ computed based on these cells has extremely low empirical power.

Then I compared $GFfit_{\perp}^{(ij)}$ with $GFfit_{\perp(\mathrm{t})}^{(ij)}$ where the slopes and intercepts in the model are generated randomly. I studied one five variables four categories case. All the slopes were generated from a uniform distribution $U(.5, 2.5)$. The three intercepts were generated from three different uniform distributions:$U(-2,-1), U(-1,1)$ and $U(1,2)$.

I first generated 1000 pseudo samples with sample size 150 from a one-factor model. The data was fitted to the correct one-factor model and the type I error rates for $GFfit_{\perp}^{(ij)}$ with $GFfit_{\perp(4)}^{(ij)}$ are shown in the following table.

TABLE 63: Type I Error Rate for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ **and** $\boldsymbol{GFfit_{\perp(4)}^{(ij)}}$, Intercepts and Slopes Generated Randomly

	Type I error rates	
(ij)	$GFfit_{\perp}^{(ij)}$	$GFfit_{\perp(4)}^{(ij)}$
(12)	0.039	0.036
(13)	0.050	0.049
(14)	0.045	0.048
(15)	0.045	0.054
(23)	0.058	0.054
(24)	0.045	0.042
(25)	0.044	0.045
(34)	0.049	0.054
(35)	0.051	0.058
(45)	0.048	0.056

We can see that only the type I error rate for $GFfit_{\perp(4)}^{(12)}$ is outside of the interval $0.05 \pm 1.96\sqrt{\frac{(0.95)(0.05)}{1000}} = (0.0365, 0.0635)$.

Then using the same parameter generating distribution, I generated 1000 pseudo samples with sample size 150 from a two-factor model. The data was fitted to the wrong one-

factor model and the theoretical and empirical power for $GFfit_{\perp}^{(ij)}$ with $GFfit_{\perp(4)}^{(ij)}$ are shown in the following table.

TABLE 64: Power for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ **and** $\boldsymbol{GFfit_{\perp(4)}^{(ij)}}$, Intercepts and Slopes Generated Randomly

	$GFfit_{\perp}^{(ij)}$		$GFfit_{\perp(4)}^{(ij)}$	
(ij)	Theoretical power	Empirical power	Theoretical power	Empirical power
(12)	0.0926	0.0892	0.0702	0.0543
(13)	0.0563	0.0564	0.0510	0.0430
(14)	**0.1446**	**0.1282**	**0.1085**	**0.1292**
(15)	0.0897	0.0871	0.0728	0.0769
(23)	0.0546	0.0471	0.0526	0.0461
(24)	0.0786	0.0830	0.0764	0.0666
(25)	0.0769	0.0953	0.0705	0.0769
(34)	0.0526	0.0605	0.0522	0.0646
(35)	0.0510	0.0492	0.0503	0.0451
(45)	0.0510	0.0594	0.0513	0.0389

From this table we can see that $GFfit_{\perp}^{(14)}$ with $GFfit_{\perp(4)}^{(14)}$ have the largest power, both theoretically and empirically. Although the theoretical power of $GFfit_{\perp}^{(14)}$ is higher than that of $GFfit_{\perp(4)}^{(14)}$, the empirical power for these two statistics are almost the same. This indicates that in this case, $GFfit_{\perp(4)}^{(14)}$ didn't improve $GFfit_{\perp}^{(14)}$.

From these simulation studies, we can see that when the subtable table is sparse, the power of $GFfit_{\perp(\mathrm{t})}^{(ij)}$ may or may not be higher than that of $GFfit_{\perp}^{(ij)}$. Even though in the situations where the two-way subtable is very sparse and $GFfit_{\perp(\mathrm{t})}^{(ij)}$ does not improve the original $GFfit_{\perp}^{(ij)}$, it did not perform worse in the simulations. And generally the empirical power of $GFfit_{\perp(\mathrm{t})}^{(ij)}$ is closer to the theoretical power than those of the $GFfit_{\perp}^{(ij)}$ when the subtable is sparse. However, if there is no sparseness problem in the subtable, generally the power of $GFfit_{\perp(\mathrm{t})}^{(ij)}$ is lower than that of $GFfit_{\perp}^{(ij)}$. Thus, I recommend to use both $GFfit_{\perp(\mathrm{t})}^{(ij)}$ and $GFfit_{\perp}^{(ij)}$ and compare them. If these two statistics result in different test results, expected frequencies could be examined.

III.3 Apply the New Method to $X_{[2]}^2$

III.3.1 $X_{[2]}^2$ Statistic

In the previous section, I have shown that when there is a sparseness problem in the subtable, $GFfit_{\perp(\mathrm{t})}^{(ij)}$ might have better performance than that of $GFfit_{\perp}^{(ij)}$. Similarly, we can apply the same idea to $X_{[2]}^2$. I will denote this new statistic $X_{[2][\mathrm{t}]}^2$, where t means computing the statistics with the t cells we choose according to the criterion introduced in the previous section.

Since $X_{[2]}^2$ is just the sum of all $GFfit_{\perp}^{(ij)}$, then we can define $X_{[2][\mathrm{t}]}^2$ as

$$X_{[2][\mathrm{t}]}^2 = \sum_{i,j} GFfit_{\perp(\mathrm{t})}^{(ij)}$$

Theoretically, if we specify the correct model, $X^2_{[2][t]}$ has a Chi-squared distribution with $df = t * C_q^2$, where q is the number of variables in the dataset. Note that the degrees of freedom equals the number of orthogonal components we used to compute $X^2_{[2][t]}$. If the model specified is wrong, then $X^2_{[2][t]}$ has a non-central Chi-squared distribution.

III.3.2 Type I Error Rate Study for $X^2_{[2]}$

To examine the performance of $X^2_{[2][t]}$. I re-examined the four-variable four-category sample size 150 Type I error rate simulation that shown at the beginning of SecIII.2. I generated 1000 pseudo datasets from a one factor model and fit it with a one factor model. The parameters for the data generating models are the following: $\alpha_{0(1)} = (-3.5, -3.5, -3.5, -3.5)', \alpha_{0(2)} = (0, 0, 0, 0)', \alpha_{0(3)} = (3.5, 3.5, 3.5, 3.5)', \alpha_1 = (1, 1, 1, 1)'$. The expected frequencies for each cells in the two-way subtables are listed below.

TABLE 65: Average Frequencies of Cells for Four-Variable Four-Category Case, n=150

Average frequencies		Category of variable i			
		1	2	3	4
Category of variable j	1	0.63	3.91	2.00	0.12
	2	3.91	35.53	26.86	2.00
	3	2.00	26.86	35.53	3.91
	4	0.12	2.00	3.91	0.63

I computed both $X^2_{[2]}$ and $X^2_{[2][t]}$. Again, since the number of categories is even, I chose four cells from each two-way subtable to compute $X^2_{[2][4]}$. The empirical Type I error rates when nominal $\alpha = 0.05$ and KS test p-values of $X^2_{[2]}$ and $X^2_{[2][4]}$ are listed below.

TABLE 66: Type I Error Rates and KS Test P-values for $\boldsymbol{X^2_{[2]}}$ and $\boldsymbol{X^2_{[2][4]}}$, Four-Variable Six-Category

	$X^2_{[2]}$	$X^2_{[2][4]}$
Type I error rate	0.1049	0.0504
KS test p-value	0.0004	0.3757

In the earlier section it was already shown that $GFfit_{\perp}^{(ij)}$ may not have a Chi-squared distribution in the sparse case. Then $X^2_{[2]}$ would not have a Chi-squared distribution since it is just the sum of all the $GFfit_{\perp}^{(ij)}$. This is demonstrated by the Type I error rate and KS test p-value shown in the table. However, we can see that $X^2_{[2][4]}$ still performs well even though the subtable is sparse.

Then I applied $X^2_{[2][t]}$ to the Type I error study for the four-variable six-category case introduced in Sec III.2.3. 500 pseudo samples were generated from a one-factor model and fitted with a one-factor model. The parameters for the data generating model are the following: $\alpha_{0(1)} = (-3.5, -3.5, -3.5, -3.5)', \alpha_{0(2)} = (-3, -3, -3, -3)', \alpha_{0(3)} = (0,0,0,0)', \alpha_{0(4)} = (3,3,3,3)', \alpha_{0(5)} = (3.5,3.5,3.5,3.5)', \alpha_1 = (2.3, 2.5, 1.9, 2.1)'$. The sample size is 1000. I conducted this simulation because this parameter setting is similar to the parameter setting I used for the power study for $X^2_{[2][t]}$ which I will show later. If the Type I error rate of $X^2_{[2][t]}$ does not look good for this parameter setting, the power study would have no meaning. Since there are six categories in this case, when computing $X^2_{[2][t]}$, I chose two different t, 4 and 16. All these cells are in the center of the subtable. The Type I error rates for $X^2_{[2]}$, $X^2_{[2][4]}$ and $X^2_{[2][16]}$ are shown in the following table. The convergence rate is 98%.

TABLE 67: Type I Error Rates for $\boldsymbol{X^2_{[2]}}$, $\boldsymbol{X^2_{[2][4]}}$ and $\boldsymbol{X^2_{[2][16]}}$, Four-Variable Six-Category

	$X^2_{[2]}$	$X^2_{[2][4]}$	$X^2_{[2][16]}$
Type I error rate	0.1183	0.0653	0.1122

It is not surprised that $X^2_{[2]}$ has a very large Type I error rate since in Sec III.2.3 I already showed that $GFfit_{\perp}^{(ij)}$ does not have a good Type I error rates. Thus $X^2_{[2]}$ would not have a good Type I error rate since it is just the sum of all the $GFfit_{\perp}^{(ij)}$. The Type I error rate for $X^2_{[2][4]}$ is within the interval $(0.0310,0.0691)$. However, when we choose 16 cells, $X^2_{[2][16]}$ has a very large Type I error rate because we may have chosen too many cells with low expected frequencies to compute $X^2_{[2][16]}$.

III.3.3 Power Study for $\boldsymbol{X^2_{[2]}}$

To examine the power of $X^2_{[2][t]}$ when the subtable is sparse, I re-examined the four variables six categories sample size 1000 power simulation shown in Section III.2.4. $X^2_{[2][4]}$ and $X^2_{[2][16]}$ were computed using the cells in the center of the two-way subtable. Both theoretical and empirical powers for $X^2_{[2]}$, $X^2_{[2][4]}$ and $X^2_{[2][16]}$ are listed below. The convergence rate is 98.2%.

TABLE 68: Theoretical Power and Empirical Power for $\boldsymbol{X^2_{[2]}}$, $\boldsymbol{X^2_{[2][4]}}$ and $\boldsymbol{X^2_{[2][16]}}$

	Theoretical Power	Empirical Power
$X^2_{[2]}$	0.1802	0.1914
$X^2_{[2][4]}$	0.3295	0.3360
$X^2_{[2][16]}$	0.1617	0.1995

So we can see that both theoretical and empirical power of $X^2_{[2]}$ are lower than those of $X^2_{[2][4]}$ when the subtable is sparse. The power of the $X^2_{[2][4]}$ may be higher because lack-of-fit located primarily in the four cells of the two-way tables where the test is focused. The theoretical power and empirical power of $X^2_{[2][16]}$ are lower than those of $X^2_{[2][4]}$. This result is consistent with the result shown in the Type I error rate study that $X^2_{[2][16]}$ does not work well for this case.

CHAPTER 4

APPLICATION, SUMMARY AND DISCUSSION

IV.1 Application

I analyzed a real data set about agoraphobia. Agoraphobia is a type of anxiety disorder in which you fear and often avoid places or situations that might cause you to panic and make you feel trapped, helpless or embarrassed. With agoraphobia, you often have a hard time feeling safe in any public places, especially where crowds gather. You may even feel unable to leave your home. This dataset consists in judgments expressed by 3305 patients about several fears. There are 5 variables in this dataset:

1.Fear of tunnels or bridges

2.Fear of being in a crowd

3.Fear transportation

4.Fear of going out of house alone

5.Fear of being alone

Each variable has three categories: “yes”, “no”, “kind of”. Our goal is to study whether these five variables can be modeled by a one-factor latent variable model. The number of all the possible response patterns is k=243. However, as most of the answers are “no”, 139 response patterns are empty. Furthermore many response patterns have a frequency less than five. The detailed frequencies are reported in Table 69.

TABLE 69: Number of Response Patterns with Small Frequencies.

Frequency	Number of Response Patterns	Number of Cases
1	46	46
2	20	40
3	11	33
4	5	20
5	2	10
>5	20	3156
Total	104	3305

We used a one-factor model to fit the data. The X^2_{PF}, $X^2_{[2]inv}$, $X^2_{[2]ss}$ and p-value are reported in Table 70.

TABLE 70: $\boldsymbol{X^2_{PF}}$, $\boldsymbol{X^2_{[2]inv}}$ $\boldsymbol{X^2_{[2]ss}}$ and P-value of the Agoraphobia Sample

	X^2_{PF}	$X^2_{[2]inv}$	$X^2_{[2]ss}$
Value	383.32	180.46	188.80
Degrees of freedom	227	40	40
P-value	<0.0001	<0.0001	<0.0001

We can see that all these three statistics are pretty large and the p-values are almost 0. This indicates that the one-factor model is not a good fit to the data.

As mentioned earlier, $X^2_{PF} = \sum_i GFfit^{(i)}_{\perp} + \sum_i \sum_j GFfit^{(ij)}_{\perp} + \sum_i \sum_j \sum_k GFfit^{(ijk)}_{\perp} + \cdots + GFfit^{(1,2,\cdots,q)}_{\perp} = X^2_{[1]} + X^2_{[2]} + \cdots + X^2_{[q]}$, we can use this sample to verify this equation. The $X^2_{[1]}$, $X^2_{[2]}$, $X^2_{[3]}$, $X^2_{[4]}$ and $X^2_{[5]}$ are reported in Table 71.

TABLE 71: $\boldsymbol{X^2_{[1]}}$, $\boldsymbol{X^2_{[2]}}$, $\boldsymbol{X^2_{[3]}}$, $\boldsymbol{X^2_{[4]}}$ and $\boldsymbol{X^2_{[5]}}$ of the Agoraphobia Sample

	Value	Degrees of freedom
$X^2_{[1]}$	24.24	10
$X^2_{[2]}$	188.80	40
$X^2_{[3]}$	102.00	80
$X^2_{[4]}$	70.26	80
$X^2_{[5]}$	22.25	32

The degrees of freedom of X^2_{PF} is 227. As we are running out of the degrees of freedom, we will omit $X^2_{[1]}$. We can see that $X^2_{[2]} + X^2_{[3]} + X^2_{[4]} + X^2_{[5]} = 383.317$, which is very close to X^2_{PF} 383.32. However, the chi-squared distribution for $X^2_{[3]}$, $X^2_{[4]}$ and $X^2_{[5]}$ may not be valid due to the sparseness in the higher order subtables.

The $GFfit^{(ij)}_{\perp}$ $M^{(ij)}_2$, X^2_{ij} and $\bar{\bar{X}}^2_{ij}$ and the p-values are shown in the following tables. I bolded the p-values less than 0.05.

TABLE 72: $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$ for the Application

(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	16.81	16.33	16.66	12.78
(13)	65.94	47.71	51.26	41.366
(14)	4.28	1.68	4.21	2.66
(15)	7.75	0.37	4.30	2.72
(23)	30.39	30.83	**31.57**	25.11
(24)	14.77	9.91	10.75	7.80
(25)	10.45	7.23	8.49	5.97
(34)	10.60	10.64	11.14	8.37
(35)	11.38	0.99	5.62	3.89
(45)	16.38	12.31	27.88	21.09

TABLE 73: P-values for $\boldsymbol{GFfit_{\perp}^{(ij)}}$ $\boldsymbol{M_2^{(ij)}}$, $\boldsymbol{X_{ij}^2}$ and $\boldsymbol{\bar{\bar{X}}_{ij}^2}$ for the Application

	P-values			
(ij)	$GFfit_{\perp}^{(ij)}$	$M_2^{(ij)}$	X_{ij}^2	$\bar{\bar{X}}_{ij}^2$
(12)	**0.0021**	**0.0003**	0.0004	**0.0017**
(13)	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
(14)	0.3690	0.4321	0.3773	0.2635
(15)	0.1011	0.8299	0.3665	0.2556
(23)	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
(24)	**0.0052**	**0.0070**	**0.0290**	**0.0202**
(25)	**0.0334**	**0.0267**	**0.0750**	0.0503
(34)	**0.0314**	**0.0048**	**0.0249**	**0.0151**
(35)	**0.0225**	0.6094	**0.2292**	0.1423
(45)	**0.0025**	**0.0021**	**<0.0001**	**<0.0001**

The results for all these statistics are consistent with each other generally. From these p-values, we can see that the association between most variable pairs cannot be explained the one-factor model.

Then I fitted the data with a two-factor model. The $X_{[2]ss}^2$ and $GFfit_{\perp}^{(ij)}$ and the corresponding p-values are shown in the following table.

TABLE 74: $\boldsymbol{X^2_{[2]ss}}$, $\boldsymbol{GFfit^{(ij)}_{\perp}}$ and the Corresponding P-values

	$X^2_{[2]ss}$	p-value
	143.94	**<0.0001**
(ij)	$GFfit^{(ij)}_{\perp}$	p-value
(12)	22.51	**0.0002**
(13)	36.19	**<0.0001**
(14)	3.34	0.5025
(15)	5.94	0.2032
(23)	20.14	**0.0004**
(24)	11.15	**0.0248**
(25)	11.12	**0.0252**
(34)	9.24	0.0552
(35)	12.23	**0.0156**
(45)	12.03	**0.0171**

From these p-values we can see that the two-factor model does not fit the data well either and the association between most variable pairs cannot be explained by this model. Since both one-factor model and two-factor model did not fit well, we may consider other models such as log-linear model.

IV.2 Summary

In summary, I studied the Type I error and power of $GFfit^{(ij)}_{\perp}$, both theoretical and empirical, and compared the performance of $GFfit^{(ij)}_{\perp}$ to that of $M_2^{(ij)}$, X^2_{ij} and $\bar{\bar{X}}^2_{ij}$. I

introduced $GFfit_{\perp(t)}^{(ij)}$ to improve the performance of $GFfit_{\perp}^{(ij)}$ when the two-way subtables are sparse and applied the improvement on $GFfit_{\perp}^{(ij)}$ to $X_{[2]}^2$.

When the correct model was fitted to the dataset and the sparseness problem was not present in the two-way subtables, $GFfit_{\perp}^{(ij)}$ and $M_2^{(ij)}$ distributed asymptotically chi-square. X_{ij}^2 does not distributed chi-square. For $\bar{\bar{X}}_{ij}^2$ if the degrees of freedom are small, its empirical distribution does not approximate chi-square well. However, when the degrees of freedom are moderate or large, $\bar{\bar{X}}_{ij}^2$ still has an asymptotic chi-square distribution.

When there is a sparseness problem in the two-way subtables, $GFfit_{\perp}^{(ij)}$ tends to have inflated type I error rate since its distribution may not be well approximated by the chi-square distribution due to the sparseness in the subtables even if the total sample size is large. In this situation, $GFfit_{\perp(t)}^{(ij)}$ may be a good remedy. Simulation results show that even though the subtable is not sparse, $GFfit_{\perp(t)}^{(ij)}$ still distributed chi-square with $df = t$.

When an incorrect model was fitted to the dataset, $X_{[2]}^2$, $GFfit_{\perp}^{(ij)}$, $M_2^{(ij)}$, X_{ij}^2 and $\bar{\bar{X}}_{ij}^2$ can be used as diagnostics to detect the source of lack of fit. If lack of fit is present in second-order marginal, then $X_{[2]}^2$ would have higher power than an omnibus statistic such as the Pearson chi-square since it represents a test that is "focused" on the second-order marginal. Similarly, if lack of fit is present in the association between variable i and variable j, then $GFfit_{\perp}^{(ij)}$, $M_2^{(ij)}$, X_{ij}^2 and $\bar{\bar{X}}_{ij}^2$ may have higher power than an omnibus statistic on the second-order marginals such as $X_{[2]}^2$. Simulation results show that $M_2^{(ij)}$ has very low power. Although X_{ij}^2 may have high power in some situations, I do not recommend to use it as diagnostics since it does not distribute chi-square theoretically. $GFfit_{\perp}^{(ij)}$ and $\bar{\bar{X}}_{ij}^2$ have the largest power among these four statistics. However, they may detect the lack-of-fit in the associations between different pairs of variables. The power of these statistics are affected by the sample size of the dataset. When the sample size decreases, both theoretical power and empirical power decrease.

When there is a sparseness problem in the two-way subtable, $GFfit_{\perp}^{(ij)}$ may have a low power. In this case, using $GFfit_{\perp(t)}^{(ij)}$ may be a good remedy to the sparseness in the

subtable. However, simulation results show that the power of $GFfit_{\perp(t)}^{(ij)}$ may or may not be higher than that of $GFfit_{\perp}^{(ij)}$. Even though in the situations where the two-way subtable is very sparse and $GFfit_{\perp(t)}^{(ij)}$ does not improve the original $GFfit_{\perp}^{(ij)}$, it did not perform worse in the simulations. And generally the empirical power of $GFfit_{\perp(t)}^{(ij)}$ is closer to the theoretical power than those of the $GFfit_{\perp}^{(ij)}$ when the subtable is sparse. However, if there is no sparseness problem in the subtable, generally the power of $GFfit_{\perp(t)}^{(ij)}$ is lower than that of $GFfit_{\perp}^{(ij)}$. Thus, I recommend to use both $GFfit_{\perp(t)}^{(ij)}$ and $GFfit_{\perp}^{(ij)}$ and compare them. If these two statistics result in different test results, expected frequencies could be examined.

When using $GFfit_{\perp(t)}^{(ij)}$, we need to decide the number of cells we choose, t, and which cells should we choose. I suggest to choose a moderate number of cells, say four or five. Simulation results suggest that we choose the cells in the center of the table. Particularly, for a dataset with c categories in each variable, if c is even, I will choose four cells corresponding to the categories pair $(\frac{c}{2}, \frac{c}{2})$, $(\frac{c}{2}, \frac{c}{2}+1)$, $(\frac{c}{2}+1, \frac{c}{2})$ and $(\frac{c}{2}+1, \frac{c}{2}+1)$. If c is odd, I will choose five cells corresponding to the categories pair $(\frac{c+1}{2}, \frac{c+1}{2})$, $(\frac{c+1}{2}-1, \frac{c+1}{2})$, $(\frac{c+1}{2}, \frac{c+1}{2}-1)$, $(\frac{c+1}{2}+1, \frac{c+1}{2})$ and $(\frac{c+1}{2}, \frac{c+1}{2}+1)$. As the two-way table becomes larger, more cells could be taken from the center of the table.

IV.3 Discussion

IV.3.1 A Method That Did Not Improve $GFfit_{\perp}^{(ij)}$

Besides $GFfit_{\perp(t)}^{(ij)}$, there is another way that might be able to improve $GFfit_{\perp(t)}^{(ij)}$. When computing $GFfit_{\perp}^{(ij)}$, we actually don't use all the c^2orthogonal components. Instead, we only use $(c-1)^2$ orthogonal components that can produce the full table. This allows us different choices of the orthogonal components. As long as the components we choose

can produce the full table, the $GFfit_{\perp}^{(ij)}$ computed should have $(c-1)^2$ degrees of freedom. But to improve the chi-square approximation, we can choose $(c-1)^2$ components corresponding to the cells that tend to have largest expected frequencies. However, simulation results show that this method may not improve the performance of $GFfit_{\perp}^{(ij)}$. There may be three reasons. Firstly, since the choice of the $(c-1)^2$ orthogonal components should be able to produce the full table, we usually cannot choose some cells with relatively large frequencies. Secondly, by choosing $(c-1)^2$ cells, it is inevitable that we might choose some cells with relatively low frequencies. Thirdly, the original $GFfit_{\perp}^{(ij)}$ already works well enough if we have to choose $(c-1)^2$ cells. For easy demonstration, I again labeled the cells in the four categories case below.

TABLE 75: Labels of Cells for Four Categories Case

Label of the cells		Category of variable i			
		1	2	3	4
Category of variable j	1	16	12	8	4
	2	15	11	7	3
	3	14	10	6	2
	4	13	9	5	1

When computing the original $GFfit_{\perp}^{(ij)}$, by default we use the $(4-1)^2 = 9$ cells on the bottom right corner of this subtable, which means cells 1, 2, 3, 5, 6, 7, 9 ,10 and 11. These cells already includes the cells in the center of this table. In Section III.2 we demonstrate that when choosing the t cells for $GFfit_{\perp(t)}^{(ij)}$, we will choose the cells in the center of the table since they will have large expected frequencies if the intercepts in the

GLLVM model are generally evenly distributed. Since when computing the original $GFfit_{\perp}^{(ij)}$ we already include all the cells in the center, choosing other cells may not be able to improve the statistic generally. To demonstrate this, I repeated the five variables five categories power study with parameter setting $\alpha_{0(1)} = (-1.59, -2.30, -1.43, -3.02, -1.26)', \alpha_{0(2)} = (-0.84, -0.38, -0.32, -1.50, -0.21)', \alpha_{0(3)} = (0.71,0.16,0.15,0.57,0.78)', \alpha_{0(3)} = (1.48,1.80,1.66,2.13,1.65)', \alpha_1 = (1.5,1.7,1.9,2.1,2.3)', \alpha_2 = (0.8,0.8,0,0,0)'$. 500 pseudo samples with sample size 300 were generated. Besides using the default cells in the bottom right corner of the subtable, I also computed $GFfit_{\perp}^{(ij)}$ using the cells in the top right corner of the subtable. The empirical power results are listed below.

TABLE 76: Empirical Power for $\boldsymbol{GFfit}_{\perp}^{(ij)}$ with Two Different Cell Selections

	Empirical Power	
(*ij*)	Default cells (bottom right corner)	Top right cells
(12)	0.1638	0.1460
(13)	0.0682	0.0520
(14)	0.0771	0.0540
(15)	0.0581	0.0460
(23)	0.0545	0.0600
(24)	0.0701	0.0460
(25)	0.0526	0.0640
(34)	0.0553	0.0560
(35)	0.0514	0.0520
(45)	0.0570	0.0400

From this table we can see that the empirical power does not differ much for these two different cells selections. So this method was not useful.

IV.3.2 Computation Time

I used R to do all the simulation studies. When the number of variables and categories increases, the computation time will increase substantially. So far, the computation time for a nine variables five categories case is about half an hour for just one sample. The main reason for such a long computation time is that we need to conduct lots of matrix production and numerical integration when computing the statistics we are investigating.

When estimating the parameters, I used two functions: the grm function in the ltm package and the mirt function in the mirt package. The grm function works significantly faster than the mirt function does. However, the grm function can only fit the one-factor model for multinomial data. The mirt function can fit the model with any number of factors. Thus, I only used mirt function for the type I error rate study for a two-factor model and used grm function for all the other simulation studies presented earlier to decrease the computation time.

Since I was doing simulation study, using parallel computing can decrease the computation time substantially. I used the foreach function in R to do the parallel computing. I run my simulation on mathpost which has 24 cores. After several experiments, I found that using 20 cores simultaneously can decrease the computation time most. For example, the computation time for the six variables four categories two-factor type I error study I presented in Sec III.1 using traditional for loop in R is about 7 hours. But using parallel computing with 20 cores reduces the computation time to about half an hour.

IV.3.3 Memory Issue

When doing my simulations, there is a memory issue when computing several very large matrices.

The first issue is about the H matrix and M matrix I introduced in Sec II.3.1. These two matrices may be the most important matrices that I need to compute since all the statistics I investigated need to be computed through these two matrices. Since the rows in the H matrix consist of a subset of the rows in the M matrix, I only stored the M matrix to save

memory. With a large number of variables and categories, the M matrix can be very large. For example, for the nine variables five categories case, the M matrix is 900 by 1953125. Using the traditional matrix generation in R, 13.1 Gb is needed to store this M matrix. Mathpost has 96 GB of memory. Since most elements in the M matrix are just zero, the M matrix is known as a sparse matrix. Using the Matrix package in R, we can generate the M matrix as a sparse matrix which save large amount of space. The M matrix for nine variables five categories case only need 812 Mb memory to store if we generate it as a sparse matrix. Although we have decreased the memory needed to store the M matrix substantially, this may still be a problem when we use parallel computing. When doing parallel computing, we use more cores to do simulations simultaneously. However, with more cores, less computing memory is allocated for each core. And the M matrix is not the only large matrix we need to store in memory. Thus, we will run out of memory very soon when we increase the number of variables and categories in our simulation. If we have to run simulations for large numbers of variables and categories, we have to use the traditional for loop in R, which is very time consuming.

Another memory issue happens when computing $GFfit_{\perp}^{(ij)}$. Recall from Section II.3.5, to compute $GFfit_{\perp}^{(ij)}$ we need to perform a regression and $GFfit_{\perp}^{(ij)}$ is the sum of corresponding sequential sum of squares. When performing that regression, we need to compute

$$\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{\mathbf{1}}{\mathbf{2}}}\mathbf{H}'$$

Where

$$\boldsymbol{D} = diag(\boldsymbol{\pi}(\boldsymbol{\theta}))$$

$$\Sigma = \Sigma(\theta) = (\boldsymbol{I} - \boldsymbol{\pi}^{\frac{1}{2}}\left(\boldsymbol{\pi}^{\frac{1}{2}}\right)' - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}')$$

Thus, $\widehat{\boldsymbol{\Sigma}}$ is a k by k matrix where k is the number of response pattern. For the nine variables five categories case, $k = 5^9 = 1953125$. Obviously, $\widehat{\boldsymbol{\Sigma}}$ will not be a sparse matrix and there is no way for us to store it in the memory. To solve this problem, we actually don't compute $\widehat{\boldsymbol{\Sigma}}$. We compute $\widehat{\boldsymbol{W}}$ directly by

$$\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{1}{2}}\mathbf{H}' = (\widehat{\boldsymbol{D}}^{\frac{1}{2}}\mathbf{H}' - \boldsymbol{\pi}^{\frac{1}{2}}\left(\left(\boldsymbol{\pi}^{\frac{1}{2}}\right)'\widehat{\boldsymbol{D}}^{\frac{1}{2}}\mathbf{H}'\right) - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\left(\boldsymbol{A}'\widehat{\boldsymbol{D}}^{\frac{1}{2}}\mathbf{H}'\right))$$

In this way, we avoid computing a matrix with extremely large dimensions. The same idea has been applied to many computations when calculating those statistics.

IV.3.4 Convergence Problem

As presented in Chapter 3, some simulations had problems of convergence. When the ML estimation algorithm for intercepts and slopes estimates does not converge, some slope estimates will be extremely large. An extremely large slope estimate will make several estimated cumulative frequencies the same value for different categories in one variable. In this case, we failed to compute the derivatives for the corresponding parameters. Without these derivatives, we cannot compute the statistics studied here. Generally, with a smaller sample size and skewness in the two-way subtable, it is more likely that the ML estimation algorithm will fail to converge. For example, I listed the convergence rate for the two-factor six variables four categories type I error rate simulations I presented in Sec III.1. below.

TABLE 77: Convergence Rate for the Two-Factor Six Variables Four Categories Type I Error Rate Simulations

	Sample size	Convergence rate
Skewed	500	.990
	150	.617
Non-Skewed	500	.995
	150	.733

We can see that when we decrease the sample size from 500 to 150, the convergence rate decreases a lot. And the convergence rate for skewed case is lower than that for Non-Skewed case.

However, even though the sample size is not small and the two-way subtable is not skewed, the convergence problem may still present. One example is the two-factor four variables three categories type I error study mentioned in Sec III.1. The sample size is 500 and the parameter setting is listed below: $\alpha_{0(1)} = (-2, -2, -2, -2)', \alpha_{0(2)} = (2,2,2,2)', \alpha_1 = (0.0, 1.0, 1.0, 0.0)', \alpha_2 = (2.0,0.1,0.2,2.0)'$. Pseudo data for 1000 samples were generated. With this setting, the dataset has neither small sample size nor skewed two-way subtable. However, the convergence rate for this simulation is 0.753. Thus I discarded this simulation.

However, if we do need to evaluate the empirical type I error rate or power of a simulation with convergence problem, we can put a cap on the slope estimates. Since the failure of computing the statistics was due to some extremely large slope estimates, putting a cap on these estimates can solve this problem. After some experiments, I found that for a two-factor model, 3.5 and -3.5 are good caps for slope estimates; for a one-

factor model, -4 and 4 are good caps for slope estimates. I repeated the two-factor four variables three categories type I error study and capped the slope estimates with -3.5 and 3.5. With this remedy, the statistics can be computed for 91% of these samples. The type I error rate of is listed in the following table.

TABLE 78: Type I Error Rate for $\boldsymbol{GFfit}_{\perp}^{(ij)}$, Two-Factor Four Variables Three Categories, Slope Capped.

$GFfit_{\perp}^{(ij)}$	Type I error rate
(12)	0.0582
(13)	0.0527
(14)	0.0648
(23)	0.0703
(24)	0.0637
(34)	0.0582

Comparing to the interval $0.05 \pm 1.96\sqrt{\frac{(0.95)(0.05)}{1000}} = (0.0365, 0.0635)$, three out of six type I error rates are outside of this interval.

Other methods of numerical estimation would have better convergence performance. Estimation by bayes methods or penalized ML would be expected to have better convergence proportions.

Even if MLE is obtained, some estimates are so extreme that calculation of the G, matrix of second derivatives fails due to $\hat{\pi}$ equals zero or one. Calculation of $X_{[2]}^2 = \boldsymbol{e}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-\mathbf{1}}\boldsymbol{e}$ may be unstable due to inverse of $\widehat{\boldsymbol{\Sigma}}_e$ matrix. The method of orthogonal component using sequential sum of square via the SWEEP operator overcomes this problem.

REFERENCES

Agresti, A. & Yang, M. C. (1987). An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics and data Analysis*, **May**, 9-21.

Bartholomew, D. J. & Knott, M. (1999). *Latent Variable Models and Factor Analysis*, Kendall's Library of statistics, London, second edition.

Bartholomew, D. J. & Leung, S.O. (2002). A goodness of fit test for sparse 2^p contingency tables. *British Journal of mathematical and Statistical Psychology*, 55, 1-15.

Cai, L., Maydeu-Olivares, A., Coffman, D.L. & Thissen, D. (2006). Limited information goodness of fit testing of item response theory models for sparse 2^p tables. *British Journal of mathematical and Statistical Psychology*, 59, 173-194.

Cagnone, S. & Mignani S. (2007). Assessing the goodness of fit of a latent variable model for ordinal data. *Metron*, LXV, 337-361.

Cochran, W. G. (1954). Some methods for strengthening the common chi-square tests. *Biomedical Journal*, **10**, 417-451.

Cox, D. R. & Hinkley D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cramer, H. (1946). *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.

Fisher, R. A. (1924). The conditions under which measures the discrepancy between observation and hypothesis. *J. Roy. Statist. Soc*, **87**, 442-450.

Goodman, L. A. (1964). Simple methods for analyzing three-factor interaction in contingency tables. *Journal of the American Statistical Association*, **59**, 319-385.

Goodnight, J. H. (1978). The sweep Operator: Its importance in Statistical Computing. SAS technical Report R-106, SAS Institute, Cary, NC.

Holst, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit test. *Biometrika*, **59**, 137-145.

Kendall, M.G. (1952). *The Advanced Theory of Statistics*, vol 1, 5^{th} ed. London: Griffin.

Koehler, K. J. (1986). Goodness-of-fit tests for log-linear models in sparse contingency tables. *Journal of the American Statistical Association*, **81**,483-493.

Koehler, K. J. & Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association*, 75, 336-344.

Lancaster, H. O. (1969). *The chi-squared distribution*. New York: Wiley.

Maydeu-Olivares, A. & Joe, H. (2005) Limited and full information estimation and goodness-of-fit testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association, 100,* 1009-1020.

Maydeu-Olivares, A. & Joe, H. (2006) Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71,* 713-732.

Mitra, S. K. (1958). On the limiting power function of the frequency chi-square test. *Annals of Statistics*, 29, 1221-1233.

Morris, C. (1975). Central limit theorems for multinomial sums. *Annals of Statistics*, **3**, 365-384.

Rayner, J. C. W. & Best, D, J. (1989). *Smooth Tests of Goodness of Fit.* Oxford: New York.

Reiser, M. & Vandenberg, M. (1994). Validity of the chi-square test in dichotomous variable factor analysis when expected frequencies are small. *British Journal of Mathematical and Statistical Psychology*, 47, 85-107.

Reiser, M. (2008). Goodness-of-fit testing using components based on marginal frequencies of multinomial data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 331-360.

Reiser, M. (2012). Limited-information statistics when the number of variables is large. Proceedings of 2012 Joint Statistical Meetings, San Diego, CA.

Reiser, M., Cagnone, S. & Zhu, J. (2014). An Extended *GFfit* Statistic Defined on Orthogonal Components of Pearson's Chi-Square. *In JSM Proceedings, Biometrics Section, Alexandria, VA: American Statistical Association.*

Zhu, J., Reiser, M. & Cagnone, S. (2015) A Power Study of the GFfit Statistic as a Lack-of-Fit Diagnostic. *JSM Proceedings, Biometrics Section, Alexandria VA, American Statistical Association.*

Tollenaar, N. & Mooijaart, A. (2003). Type I errors and power of the parametric bootstrap goodness-of-fit test: Full and limited information. *British Journal of Mathematical and Statistical Psychology*, 56, 271-288.

Wittchen, H. & Gloster A. T. & Beesdo-Baum, K. & Fava, G. A. & Craske, M. G. (2010). Agoraphobia: a review of the diagnostic classificatory position and criteria. *Depression and Anxiety,* 27, 113-133.